AD ________________

MIPR NO: 95MM5560

TITLE: Vulnerability of Female Produced Speech in Operational Environments

PRINCIPAL INVESTIGATOR: Dr. Charles W. Nixon , Linda J. Morris, Angela R. McCavitt, Mark P. McDaniel, Timothy R. Anderson, Donald G. Yeager, Richard L. McKinley

CONTRACTING ORGANIZATION: Armstrong Laboratory, AL/CFPR
Wright-Patterson AFB, OH 45433-7008

REPORT DATE: April 1996

TYPE OF REPORT: Final

PREPARED FOR: U. S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. AGENCY USE ONLY** *(Leave blank)*

**2. REPORT DATE**
April 1996

**3. REPORT TYPE AND DATES COVERED**
Final - 5 Dec 94 to 30 Sep 95

**4. TITLE AND SUBTITLE**
Vulnerability of Female Produced Speech in Operational Environments

**5. FUNDING NUMBERS**
95MM5560

**6. AUTHOR(S)**
Dr. Charles W. Nixon, Linda J. Morris, Angela R. McCavitt, Mark P. McDaniel, Timothy R. Anderson, Donald G. Yeager, Richard L. McKinley

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Armstrong Laboratory, AL/CFPR
Wright-Patterson AFB, OH 45433-7008

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
US Army Medical Research and Mâteriel Command
Fort Detrick, MD 21702-5012

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution is unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

The role of women in combat environments has expanded dramatically. The accommodation of females in these previously male dominated environments is vital to fully successful mission accomplishment. Effective voice communications are critical to successful strategic and tactical operations. Current aircraft communications systems and components were optimized for male voice characteristics. Based on laboratory experience and knowledge of the basic characteristics of male and female acoustic speech, many questions exist concerning possible differential effects of operational variables on female speech. This study investigates the impact of the following factors on the perception of female speech: (a) the different spectra of operational noise environments, (b) the response characteristics of noise-cancelling microphones, (c) digital encoding and decoding, and (d) accuracy of automatic speech recognition systems.

**14. SUBJECT TERMS**
Female Speech, Speech Perception, Female vs Male Speech, Speech in Noise, Speech Intelligibility, Voice Communications
Defense Women's Health Research Program

**15. NUMBER OF PAGES**
76

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UNLIMITED |

## GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

**Block 1.** Agency Use Only *(Leave blank)*.

**Block 2.** Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

**Block 3.** Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

**Block 4.** Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

**Block 5.** Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

| | | | |
|---|---|---|---|
| **C** | - Contract | **PR** | - Project |
| **G** | - Grant | **TA** | - Task |
| **PE** | - Program Element | **WU** | - Work Unit Accession No. |

**Block 6.** Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

**Block 7.** Performing Organization Name(s) and Address(es). Self-explanatory.

**Block 8.** Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

**Block 9.** Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

**Block 10.** Sponsoring/Monitoring Agency Report Number. *(If known)*

**Block 11.** Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

**Block 12a.** Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

**Block 12b.** Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

**Block 13.** Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

**Block 14.** Subject Terms. Keywords or phrases identifying major subjects in the report.

**Block 15.** Number of Pages. Enter the total number of pages.

**Block 16.** Price Code. Enter appropriate price code *(NTIS only)*.

**Blocks 17. - 19.** Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

**Block 20.** Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_X_ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Charles W Nixon 3/22/96

PI - Signature Date

# TABLE OF CONTENTS

**FINAL REPORT** **1 APRIL 1996**

# Vulnerability of Female Produced Speech in Operational Environments

**"No other essential activity in aircraft operations is as vulnerable to failure through human error and performance limitations as spoken communications."** Monan (1986) (21)

## INTRODUCTION

This research program examined the perception of female speech produced in operational environments by listeners in operational environments. Emphasis was on female aviators and selected systems and conditions that are elements of typical military aircraft voice communication systems. Speech performance was measured in the cockpit noise environments of four different types of aircraft, with noise-cancelling microphones, with digital speech coders and decoders, and automatic speech recognition systems (voice controllers). Perception of female speech was evaluated relative to male speech perception and to performance criteria that indicate the relative effectiveness of the female speech under operational conditions.

Vigilance is essential to ensure effective voice communications critical to successful strategic and tactical operations. Numerous system, operator, and environmental factors can degrade effective communications to marginal or unacceptable levels. The basic designs and the performance of current aircraft audio communication systems have remained unchanged for several decades and need to be upgraded to incorporate current technologies. Special speech vocoders and encryptors dismantle and later reconstruct the acoustic speech signal that is often less robust and more vulnerable to noise than the original signal. Noise can directly degrade speech communications by interfering with or masking the speech signal. It can further indirectly degrade the signal by causing temporary and permanent noise induced hearing loss in the aviators. Noise can also interfere with the operation of voice recognition or voice control systems which are unable to extract the aviator speech signal commands from the noise. These factors have been dealt with for a long time without full success. They must receive continual attention to maintain effective voice communications and avert difficult and life threatening operational situations caused by the inability to communicate.

A situation is emerging that introduces a new factor that may, or may not, decrease the effectiveness of voice communications. Women are already flying high performance aircraft and their increasing presence in the cockpits and crew stations of Department of Defense (DoD) strategic and tactical aircraft is assured. Current aircraft audio communication systems and components were optimized for male voice characteristics and may not fully accommodate the female voice. Current knowledge of the perception of female speech, particularly in the harsh

environments of military aviation, is not sufficient to allow reliable estimates of female speech performance in the cockpit environment. This project studied the information necessary to identify significant differences, if present, in the perception of female and male speech. Differences that would prevent female speech from communicating effectively in current weapon systems were addressed. Difficulties with the perception of female speech would affect all aviators.

**Fundamentals of Human Speech**

In the study of the human voice, the variability in characteristics from talker to talker is a dominant feature. Consequently, when the acoustic speech signals of talkers are analyzed, different acoustic spectra are obtained. However, a basic feature of the speech sounds and the frequency regions in which their maximum amplitudes occur is that they are about the same and are generally independent of the talker. It is this basic feature that allows the acoustic characteristics of speech to be studied systematically.

The perception of female and male speech is essentially equivalent under almost all typical living conditions (ranging from a whisper in church to a shout at the playground); however, recognizable gender differences are obvious. The bases for these differences are associated with the acoustic speech signals generated by the male and by the female talker. The acoustic components of the female speech signal are almost always higher in frequency than those of the male. The fundamental frequency of the average female voice is about 250 Hz and of the average male voice is about 125 Hz. The speech spectra for average male and female speech are similar, with the female spectrum higher than the male spectrum by about 5 to 10 dB above 4000 Hz and lower by about 12 dB below 125 Hz. In the female voice, the high frequencies of the vowel sounds are 5 to 15 percent higher, the mid-high frequencies 5 to 25 percent higher, and the low frequencies up to 35 percent higher than the corresponding frequencies in the male voice. The average speech power for males is 34 microwatts and for females is 18 microwatts which corresponds to a difference of about 3 dB at conversational speech level (8).

In addition to gender differences, the acoustic features of an individual's speech are continuously changing for various voluntary and involuntary reasons. A talker may emphasize segments of speech, alter speech rate and level, shout, talk during physical exertion, and speak with emotion. Speaking in a raised voice, in order to be understood in the presence of a background noise or to talk to a distant listener, requires increased vocal effort. The accompanying muscle strain usually causes an increase in the pitch of the voice, and can cause vocal cord fatigue over time. These changes also influence the differences between female and male speech.

Human speech is very robust and is easily understood in many distorted forms. Accents, incorrect pronunciation, foreign dialect, speech compression, peak clipping, and digital coding and decoding of the speech signals may sound unnatural, yet be very intelligible. In spite of the robust nature of speech and its ability to be universally understood, it is subject to degradation under various conditions. Degradation can be caused by unfavorable speech-to-noise ratios, distortions, communications channels, terminal equipments that include microphones and earphones,

workload, stress, and the individual talker and/or listener. Factors which degrade speech communications in military applications must be identified and their impact on operations evaluated.

## Operational Variables and Communications Effectiveness

Crew stations in military aircraft contain many factors with the potential to reduce voice communications effectiveness even though the stations have been designed to optimize performance. Perhaps the most pervasive factor at these stations is acoustic noise. Noise is caused by numerous sources including vehicle propulsion systems, environmental systems, life support systems, weapons fire, and air turbulence, as well as the voice communication system itself. One of the primary effects of the noise is masking of the voice communication signals. In general, when the level of the noise in the frequency region of the speech sounds exceeds the level of the speech, communications are degraded. The ratio of the level of the speech to the level of the noise (signal-to-noise ratio, SNR) provides an estimate of the level of the speech performance; the higher the ratio, the better the speech performance. Also, some learning is involved in becoming an effective communicator in noise environments; understanding speech in noise improves with practice (16). Persons experienced with communicating over military systems in noise usually perform very well.

### *Noise*

Over the past three decades, human-in-the-loop voice communications research has been conducted in the Bioacoustics and Biocommunications Branch of the Air Force Armstrong Laboratory. Each of the major communication research facilities within the branch contains ten communication stations; consequently the standard procedure used in investigations is to simultaneously utilize ten experimental subjects. The panels of trained subjects, over the three decades of research, have consisted of five males and five females. Although research during this period did not focus on female speech, some studies involved comparisons of female-male speech performance. In general, these measurements and observations have revealed that the performance of female speech has, in most instances, been lower or less effective than male speech under the same conditions. In some situations the speech of both genders is acceptable even though the female speech is less intelligible. In other situations only the female speech is unacceptable. The current study is concerned with the systematic measurement and evaluation of some of these differences and their importance in selected environments.

In a 1991 summary of a study by Backs and Walrath, it was stated that "...under conditions of high noise stress, female speakers were less intelligible than males..." (3). In an earlier study of voice communications in simulated cockpit noise, a systematic difference was measured between the intelligibility of male and female talkers (15). In levels of noise at 95 dB and below, there was essentially no difference in intelligibility. At levels of noise at 105 dB the female talkers were seven percent less intelligible than males and at 115 dB the difference increased to ten percent. The differences at both the 105 dB and 115 dB levels of noise were significant ($p < 0.05$).

Additionally, a study of positive pressure breathing effects on speech intelligibility was conducted in aircraft noise at levels of 65 dB, 95 dB, 105 dB, and 115 dB (20). The results of this study reflect the same general observations of effects of noise on female and male speech intelligibility. The differences between the perception of female and male speech increase as the levels of the noise increase, with the female speech becoming less intelligible. The maximum decreases of female speech intelligibility occur at the highest levels of noise. In that study, only the differences at the 115 dB level of noise were significant ($p < 0.05$).

### *Noise-Cancelling Microphones*

The adverse effect of noise on speech transmission promoted the development of noise-cancelling microphones, again based on the male voice. Aviators now use two general types of noise-cancelling microphones, a "kiss-to-talk" microphone (lips touch microphone for maximum performance) for applications such as oxygen masks, and a boom type microphone for headsets and helmets worn by personnel in environs such as tankers and transport aircraft. The speech intelligibility of female aviators using either type of microphone was not measured prior to this study. An early evaluation of female and male speech intelligibility was conducted with the M-101 microphone, a former Air Force standard microphone. The M-101 microphone was compared to a modified M-101, reduced 50 percent in thickness to improve its fit in an oxygen mask. The intelligibility of the female and male speech measured in a 95 dB level of noise was essentially the same for the standard M-101 microphone. However, the speech intelligibility of the male voice increased eight percent with the "thin M-101" microphone; whereas, the female speech intelligibility increased only three percent. While these differences were not statistically significant, female speech intelligibility was lower than the male speech intelligibility as is usually observed under noise conditions (14).

### *Speech Coders*

Speech coders (vocoders) have been added to military voice communication systems to increase and maintain the reliable transfer of information. Speech coders convert the analog speech signal to digital units which are transmitted to the receiving station where they are converted back to speech. During this process some of the analog speech signal is lost; the amount of the signal that is lost is a major factor determining the quality of the coded speech. The effectiveness of the vocoding process and the amount of information lost depends on the characteristics of the analog-to-digital-to-analog conversion system.

Earlier research with three versions of the DoD standard Linear Predictive Coding (LPC-10) speech coder demonstrated that its intelligibility was poor and that it was vulnerable to voice communication degradation due to acoustic noise at the listener (17). These data were revisited as part of the current study and performance of the male and female speech was extracted and examined. Female speech in high performance aircraft and in combat was not an issue at the time of the original study. Although the sample size was very small (two female and three male talkers), the average intelligibility with the three LPC-10 vocoders at four levels of

noise was essentially the same for males and females. On the basis of other research efforts, involving four levels of noise, it was predicted that the female speech would be less intelligible rather than equal to the male speech. Consequently, when the study and the instrumentation were re-examined, it was discovered that the gain of the speech signal available to the subjects (who are usually able to individually adjust the gain for their own headset systems) was limited. This undiscovered limitation prohibited the subject from increasing the gain of her/his individual intercommunication system to improve speech communications. Without limitation on gain, it is assumed male speech perception would have been better than female speech perception. Since whether or not the difference would be significant cannot be estimated from the available information, the intelligibility of female speech processed by the standard LPC-10 vocoder and perceived in noise environments must be determined empirically.

*Automatic Speech Recognition (ASR)*

Automatic speech recognition or voice control systems are very effective when properly trained to recognize the talker and when used in relatively quiet environments. However, the success of these systems has generally been limited in high level noise environments because of their inability to discriminate the components of the acoustic noise signal from the acoustic speech signal. Even though the speech recognition system has been taught to recognize a talker ("memorizes" speech components during training), it can be fooled and will interpret components of the noise as elements of the speech, resulting in incorrect recognition. The aircraft cockpit is a particularly hostile environment for voice control systems, yet it is one that can derive substantial benefit from the successful implementation of voice control.

Despite significant advances in microphone technology, a microphone's sensitivity to noise and non-speech sound complicates the operation of automatic speech recognition systems. Current noise-cancelling microphones reduce the level of the noise as a function of frequency, but they do not eliminate the noise. Also, the acoustics inside the oxygen mask are further complicated by non-speech sounds such as the aviator breathing noise, as well as valve noise during each respiration cycle, added to the external noise that has reached the inside of the oxygen mask (18). In spite of these persistent problems, state-of-the-art speaker-dependent and speaker-independent voice control systems have been designed specifically for the cockpit environment (23). Some of the manufacturers of these systems report word recognition accuracy of over 80 percent for connected digits and over 95 percent for words spoken as two-word phrases in 90 dB of noise (90 dB is well below many operational noise environments). Speaker-dependent systems generally function with limited vocabularies and with substantial talker training. Speaker-independent systems do not require training. Recognition accuracy also varies with the talker.

Voice control technology is already present, to a limited degree, in several aircraft. Utilization of voice control in the noisy cockpit is expected to increase; however, no major breakthroughs in voice control technology appear to be on the horizon. There is no database of the recognition of female speech by voice control systems in cockpit-like noise environments. Knowledge of factors such as the lower acoustic power of the female voice and its

reduced intelligibility in higher noise levels indicates that voice control with the female voice in operational noise environments must be evaluated.

*Voice Warning*

An indirectly related area of female speech perception is that of voice advisories and voice warning signals. The initial installation of a voice warning system in an Air Force military aircraft was in 1961 when an audio tape system was installed in the fleet of B-58 Hustler aircraft. Early evaluations indicated that aviators felt that voice warnings contributed to flight safety, that pilot reaction time was improved, that warning recognition time improved by six to nine seconds, and that the female voice was the preferred warning signal (9). Voice warning systems have been evaluated in terms of aviator preferences (includes voice quality) and of quantitative metrics such as accuracy of response, reaction time, and speech intelligibility. Although aviators are relatively firm in their initial judgments for particular voice characteristics, their preferences tend to change with their continued exposure to those voices in operational situations (24).

Contrary to early beliefs, subsequent research has demonstrated that the female voice is not the preferred warning signal, and it usually ranks low in terms of both preference and quantitative metrics. Reportedly, the male voice has greater accuracy and a shorter response time than the female voice in a 105 dB noise environment (9). Another report measured no differences in the intelligibility of male and female voice warnings in a noise environment of 95 dB (24). Moreover, a distinctive, mechanical quality voice (possibly synthesized) that could be recognized in a background of human voices was preferred over either female or male voice warnings. New technologies provide a great deal of flexibility, and at relatively low cost, for the generation of voice warning systems. It is not unreasonable to expect that successful systems might use a variety of voices, both human and synthesized, to create a menu driven system allowing the aviator to select the suite of voice and auditory warning signals that she/he believes will provide the best performance. It is unclear if there is any relationship between the low ranking of female speech as a voice warning signal and its "lower than male speech" intelligibility under various conditions such as high levels of noise.

## RESEARCH OBJECTIVES

Dramatic transitions are underway with the acceptance of females as military aviators in a profession formerly occupied only by males. The total aviation environment and all related facilities and equipments were designed and evaluated for the male. Numerous efforts are underway attempting to identify those situations in the aviation environment with which the female is not fully compatible and to evaluate their impact on performance and safety. Voice communication and its effectiveness under a variety of different operational situations and circumstances is one of the most important areas under investigation. It is almost universally accepted that in-flight voice communications must be free of errors. In a report on civil aviation, Billings and Cheany (1981) state, "Problems in the transfer of information between the aviation system were noted in over 70 percent of 28,000 reports submitted by pilots and air traffic controllers...during a 5-year period 1976-1981. These problems are related primarily to voice

communications..." (21) It would be unusual to find a reader who does not know of some situation in which a breakdown of communication has resulted in an unacceptable consequence.

Under normal conditions, the understanding of female and male speech is equivalent even though there are obvious differences in the acoustic speech signals. In situations where factors such as noise degrade speech, female speech intelligibility is reduced more than that of males. These differences in speech associated with gender, and the reduction in intelligibility, tend to increase with increasing levels of noise. When this reduction in intelligibility reaches certain levels, speech communication is no longer effective.

The research objectives of this study are to quantify the differences between the perception of female and male speech relative to those factors in operational situations that influence voice communications, to determine whether the reductions in speech performance are or are not significant relative to operational environments, and to propose actions to minimize significant effects, where feasible.

The specific questions selected for investigation are, to what extent is the perception of female (and male) speech affected by:

(a) the different cockpit noise environments (spectra) of four operational aircraft,
(b) the response characteristics of standard military noise-cancelling microphones,
(c) digital encoding and decoding of the speech signals with the DoD standard LPC-10 and Continuously Variable Slope Delta (CVSD) vocoder, and
(d) voice controllers/automatic speech recognition systems.

## APPROACH

The acoustic speech signal differences based on gender have not been systematically investigated in environments emulating operational conditions that include standard communication systems and equipments in realistic acoustic environments. This study initiates such an effort; however, the large number of these environments and the time required to emulate all of the operational conditions of interest are prohibitive for a one-year study. The research team considered a variety of questions relative to their possible impact on the mission, time frame of the study, and laboratory resources that could immediately be brought to bear on the issues. It was agreed that the four proposed phases of the study would evaluate communication performance in a reasonable representation of operational conditions and speech communication technologies.

The initial phases of the study examined speech performance in typical aircraft cockpit noises (5, 6, 7, 10). Four different aircraft noise spectra were selected to represent the range of cockpit noise environments in which female aviators are found. These cockpit noise environments include the low frequency spectra of the C-130E aircraft and MH-53 helicopter, the relatively flat spectrum (up to 4000 Hz) of the C-141B, and the higher frequency spectrum of the F-15A tactical aircraft. The noise spectra shown in Figure 1 represent the levels of noise experienced in the fixed-wing aircraft cockpit positions during normal cruise flight conditions and during hover at 50 feet in the helicopter aircraft. The flight deck, as well as other crew locations, can experience

levels of noise much higher than those observed during cruise and hover. In Phase I, speech performance was measured for each of the aircraft in four different levels of the cockpit noise spectra. In Phase II, the relative effectiveness of the current standard noise-cancelling microphones was examined in the same noise environments employed in Phase I.

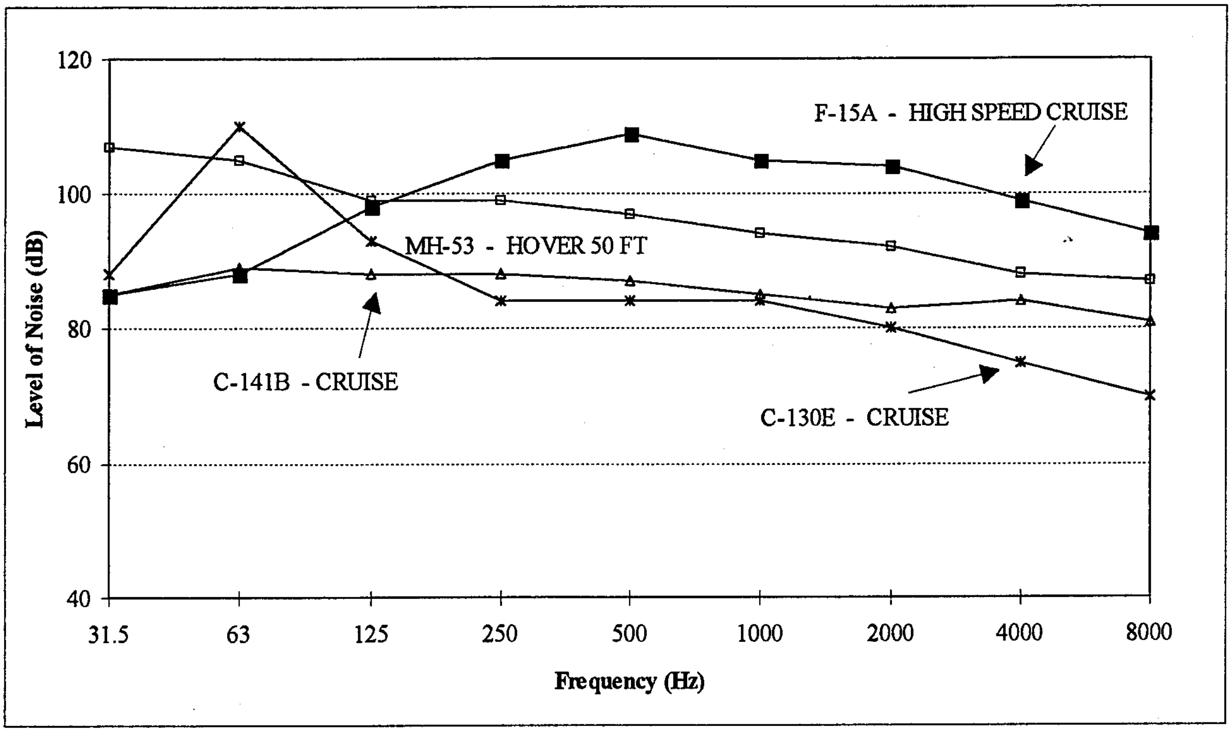



*Figure 1: Aircraft cockpit noise spectra.*

The intelligibility of the male and female speech processed by the DoD standard LPC-10 speech coder and a high quality speech coder (Continuously Variable Slope Delta modulation system, CVSD) was examined in Phase III. As noted earlier, the coder converts the analog speech signal to a digital signal that is transmitted to the receiver where it is reconverted to speech. Some of the speech signal is lost in this conversion process. Phase III examined the robustness of the reconstructed female speech in the presence of the four aircraft noise conditions of Phase I.

Control of critical operations in the cockpit by voice commands requires highly accurate recognition systems. In Phase IV, the recognition accuracy of female and male speech by two different automatic speech recognition (ASR) systems was evaluated in two cockpit-noise environments. Voice control is already present in cockpits, and eventually it is expected to extend to more aircraft and require greater numbers of commands per aircraft. Some of the better ASR systems are reported to obtain 90 to 95 percent word recognition accuracy in noise levels of about 90 dB. However, accuracy can fall off sharply as the level of the cockpit noise increases. Recognition accuracy by ASR systems of male and female speech in aircraft noise has not been previously reported. The difference between female and male voice control was empirically examined in Phase IV of this study.

### Criterion Measure

The criterion measure for Phases I, II, and III is the percent correct intelligibility of the Modified Rhyme Test (MRT) (10). The MRT is the test of choice for evaluating the performance of military communication systems and equipments. The materials consist of word lists that are equivalent in intelligibility. Each list contains 50 monosyllable words in the form of consonant-vowel-consonant. During the investigation, the talker speaks each of the 50 test words in a list in the carrier phrase, "Number ___, you will mark ___ please." The listeners select the word they believe was spoken by the talker from a set of six words that rhyme with the spoken word (Appendix C). The listener's intelligibility score is the percent correct adjusted for correct answers obtained by guessing (2.4 × number correct - 20). The score for the experimental condition is the average of the scores of the ten listeners. The MRT does not require extensive training of subjects and is relatively simple to administer, score, and evaluate. The measurement of speech intelligibility in this study was accomplished in accordance with the American National Standard, S3.2-1989, Method for Measuring the Intelligibility of Speech Over Communication Systems (2).

The criterion measure for Phase IV is recognition accuracy of words and sentences by ASR systems. The test procedure does not require human listeners to respond to the speech. The talkers spoke the sentences to the ITT ASR system live while being simultaneously recorded on high quality digital tapes and being scored by the computer. The recorded sentences of the talkers were then presented to the IBM ASR system and recognition accuracies were calculated by the computer.

### Performance Criteria

The Bioacoustics and Biocommunications Branch, at Wright-Patterson Air Force Base, Ohio, maintains a vigorous research program in all aspects of voice communications effectiveness. The laboratory uses dedicated facilities designed to evaluate all the system, operator, and environmental variables that can degrade voice communications. The data and experiences obtained using the Modified Rhyme Test, the standardized procedures, and the Voice Communication Research and Evaluation System (VOCRES) laboratory facilities (Figures 2 and 3) revealed a high relationship with performance in the operational situation. For example, a head-mounted bone conduction microphone, designed for Air/Sea Rescue applications, exhibited performance that failed the laboratory performance criteria. Development continued, but the microphone subsequently failed the Operational Test and Evaluation program. A different microphone used in a new low-profile oxygen mask also failed the speech communications performance criteria, but was still provided to operational fighter pilots. Field performance was so poor that the aviators were prohibited from flying with that microphone. Conversely, active noise reduction headsets, crew helmets, and new noise-cancelling microphones are examples of equipments that were acceptable under the performance criteria and remain highly successful in the operational situation. These examples verify the relationship between the Biocommunications Laboratory performance criteria and actual performance under operational conditions. Consequently, a set of speech intelligibility performance criteria based on data measured in the laboratory and subsequently confirmed in operational environs cited above, was

adopted several years ago and it continues to be utilized to successfully estimate and predict corresponding performance in the field.

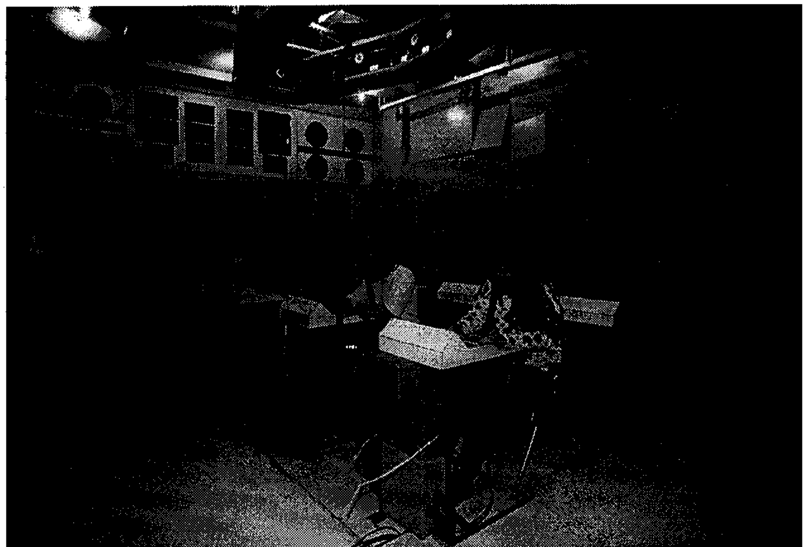


*Figure 2: Voice Communications Research and Evaluation System (VOCRES).*

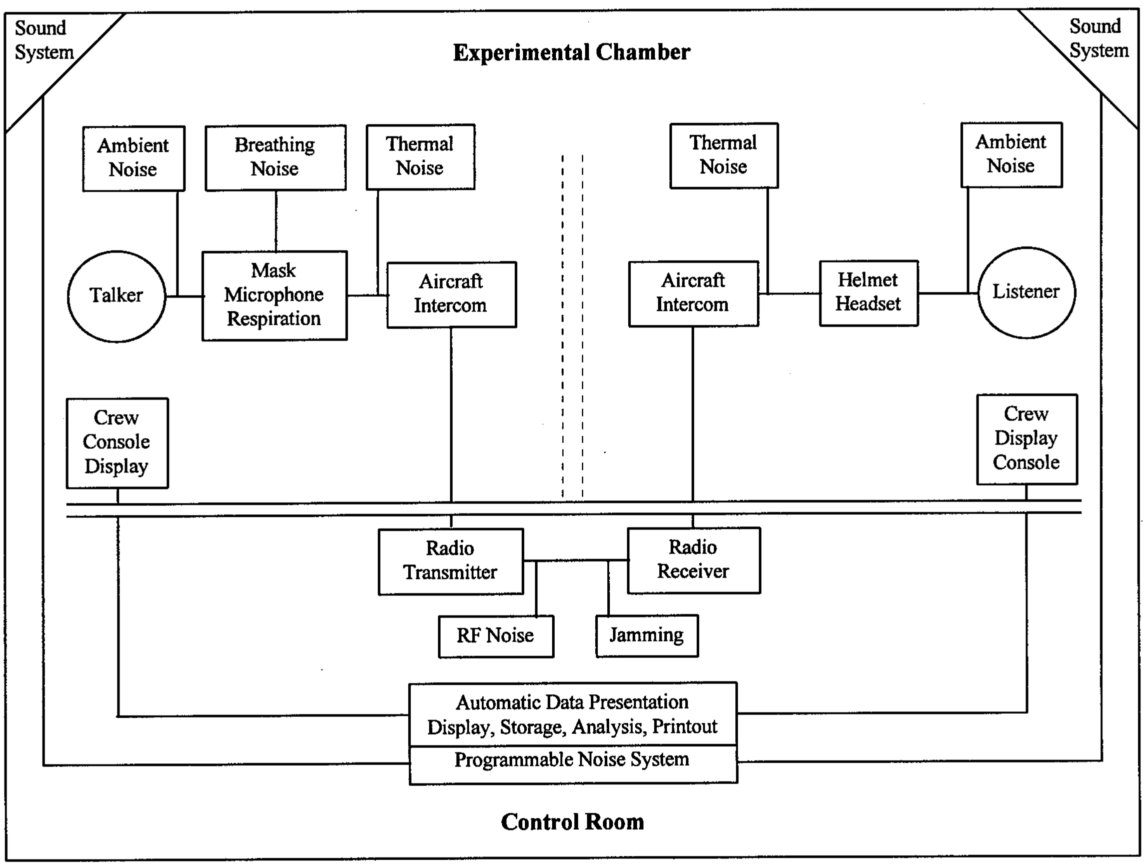



*Figure 3: Configuration of the VOCRES facility.*

The performance criteria predict systems, components, and materials displaying speech intelligibility performance below 70 percent correct (MRT) are typically unacceptable in corresponding operational applications. Those systems and components with performance in the range from about 70 percent to 80 percent are considered marginal and their success in the field depends on the specific conditions under which they are utilized. Systems with marginal performance may be used in situations where there is ample time to repeat messages to achieve understanding. Those exhibiting intelligibility performance of about 80 percent correct and above are fully acceptable under operational conditions. Speech performance measured under the various conditions in this study (Phases I-III) was examined in terms of these performance guidelines. These guidelines have been very useful in many situations, such as those in which differences in the measured speech intelligibility are statistically significant but the amount of the difference is so small that it is not meaningful in field situations. These criteria are valid for evaluations accomplished utilizing the facilities and procedures in the Armstrong Laboratory, Biocommunications Laboratory.

Phase IV data could not use the above performance criteria because the criteria were developed for use with MRT data collected in VOCRES facility and/or the Performance and Communication Research and Technology (PACRAT) facility. Performance criteria do not exist for use with the data in Phase IV since no operational performance data have been collected using automatic speech recognition (ASR) systems. Performance criteria need to be developed for use with ASR systems but it must be remembered that these criteria would be ASR, task, and vocabulary specific.

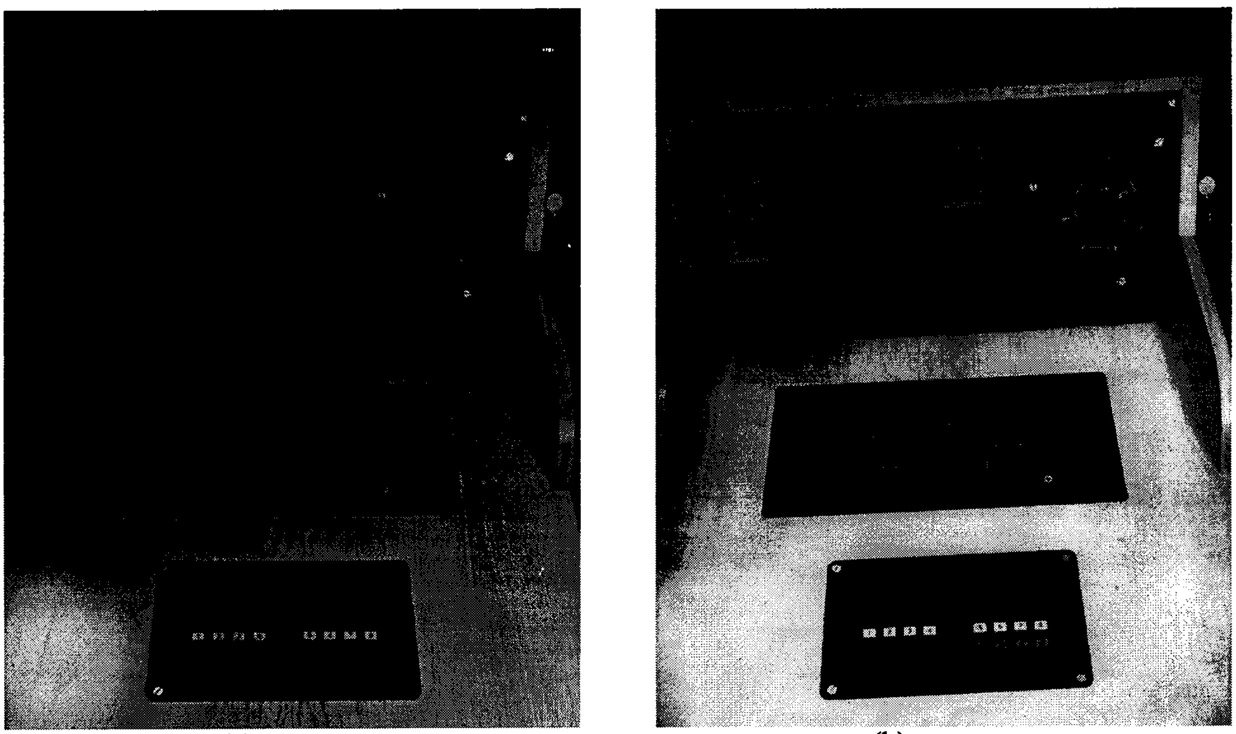
## Subjects

This investigation utilized human subjects who were experienced in voice communications research. Subjects were recruited from the general population and were paid an hourly rate for their participation. All spoke midwestern American English and none exhibited a noticeable accent, dialect, or speech problem. Twenty adult subjects, ten males and ten females, participated throughout all phases of the study. All subjects participated as talkers and a subset of ten subjects (five male and five female) comprised the listening panel. Subjects exhibited normal hearing sensitivity and middle ear function, as verified by pure tone audiometry and tympanometry, prior to participation in the study. Noise exposures were maintained within the daily exposures allowed by Air Force regulation and monitoring audiometry was performed biweekly throughout the study to insure no individuals incurred a hearing threshold shift. A communication headset/helmet was custom fit to each subject and worn by that subject throughout the study. Sound attenuation of each headset/helmet (Appendix B) was measured while worn by each subject to insure that she/he received adequate hearing protection during the study (1).

## Facilities and Equipment

Phases I, II, and III of the study were conducted in the VOCRES facility in the Armstrong Laboratory Crew Systems Directorate (13). This voice communication research system located in a large reverberation chamber contains the operator, system, and environmental variables known

to most directly affect voice communication effectiveness (Figure 2). VOCRES consists of a central processing unit that controls the experimental sessions and the subject stations (Figure 3). The facility contains ten individual automated communication stations which provide simultaneous measurement of all test subjects. Each station is equipped with an alphanumeric light emitting diode (LED) display, a subject response unit consisting of special keyboards for entering performance responses to the central processing unit, and a large volume unit (VU) meter that indicates voice level of the speech produced by the talker at that station (Figure 4). The stations contain Air Force standard helmet/headsets, air respiration systems with oxygen masks, and aircraft intercommunication systems. Aircraft radios, electronic warfare instrumentation, secure speech units, speech vocoders, and a wide-passband research intercommunication system are also imbedded in the VOCRES. In Phases I and II, an additional communication station was located inside VOCRES to accommodate the individual talker in the same noise environment as the ten-member listening panel.



(a) (b)

*Figure 4: (a) VOCRES talker station. (b) VOCRES listener station.*

VOCRES also contains a programmable sound system that can generate high intensity levels of noise in the laboratory. The overall system allows the accurate recreation in the laboratory of essentially any operational voice communication situation and noise environment. Air Force standard headsets, helmets, and microphones used for this study are those currently found in operational aircraft. The headset/helmet systems are listed with the appropriate aircraft in Table 1.

| Aircraft | Headset/Helmet System | Microphone |
|---|---|---|
| C-130E | H-157 Headset | M-87 |
| C-141B | H-157 Headset | M-87 |
| F-15A | HGU-55/P Helmet with MBU/P oxygen mask | M-169 |
| MH-53 | SPH-4AF Helmet | M-87 |

*Table 1: Aircraft, headset/helmet, and microphone combinations used in Phases I and III*

Two digital speech coding systems, called vocoders, were selected to process the speech signals in Phase III. These systems segment, process, and code the natural speech signal, and later decode it to provide the speech output. The vocoders utilized in this study are the DoD standard Linear Predictive Coding (LPC-10) and the Continuously Variable Slope Delta (CVSD) modulation speech coding systems. LPC-10 predicts the current speech sample from a linear combination of previous speech samples. It is based on the voicing, pitch, reflections, and amplitude of the speech. This information is processed into the standard LPC format. LPC is reportedly vulnerable to noise (17,18). CVSD uses an algorithm that codes only the difference between one speech sample and the next sample. Basically, the difference is coded and used to predict the next speech sample in this ongoing process. CVSD is robust in noise (19).

The two automatic speech recognition (ASR) systems employed in Phase IV are the ITT VRS-1290 and the IBM VoiceType. These systems represent two different technologies for continuous speech recognition. The ITT VRS-1290 is a speaker-dependent ASR system. Each individual talker must train the recognition system to recognize her/his speech production. This is accomplished by the talker speaking each vocabulary word and sentence into the system a number of times. The ITT system has a vocabulary of 500 words and uses the Dynamic Time Warping (DTW) technology to perform its pattern recognition and matching at the word level. This system uses special purpose hardware and a personal computer (PC). An earlier version of this system was flown in an F-15A aircraft during the mid 1980s (25). The current version of this system has been tested and flown in an Army helicopter (11). The IBM VoiceType is a speaker-independent system. It requires no specific training of the system to recognize the individual talker. The IBM system has a vocabulary of over 1,000 words and uses the Hidden Markov Model (HMM) technology to represent words with sub-word units called phonemes. This process enables additional words to be added to the system recognition vocabulary by adding the sequence of phonemes for the new word to the dictionary. The IBM system runs on a PC using an analog-to-digital converter. This system has been evaluated in laboratory environments (24).

## Experimental Systems Calibration and Measurement

Prior to data collection, all equipment was calibrated to ensure reliability, conformity to specifications, and accuracy. Earphone outputs were measured for the H-157 headset, and for the HGU-55/P and SPH-4AF helmet communication units. Each earcup was placed on an artificial ear with a flat plate coupler and 2 volts rms were applied at frequencies of 125, 250, 500, 1k, 2k,

4k, and 8k Hz. Output values were logged and compared; differences between the outputs of the two earphones in a headset unit did not exceed 5 dB. Frequency responses were obtained from measurements of the voltage output of each M-87, M-162, and M-169 noise-cancelling microphone by placing the microphone 1/4" away from an artificial voice with an output level of 95 dB using a Brüel and Kjær 4134 reference microphone (Appendix A). One microphone of each type, representative of the measured average response of that type of microphone, was selected for use in the experiment.

The Voice Communications Research and Evaluation System (VOCRES) was calibrated by passing eight pure tones at octave spacings from 100 Hz to 6300 Hz through the system for analyses by an audio analyzer. The speech calibration frequency was 1000 Hz. Distortion and acoustic noise at the headset of each station were within specifications, background noise was minimized, and VU meters were adjusted to provide appropriate visual feedback of voice volume to the talker at each station. Each of the ten stations was characterized by collecting frequency response data for the headphone and microphone.

## GENERAL PROCEDURES

### Phases I, II, and III

All data were collected with both the talker and listeners in the same noise environments. As previously noted, the experimental design required the measurement of the perception of the speech of twenty talkers by a panel of ten listeners. Twenty talkers (usual procedures only use ten talkers) were selected to expand the applicability of the data and findings of the study. Experience with voice communications in noise environments has revealed greater variance among the speech of groups of talkers than among listeners (18).

The C-130E, C-141B, F-15A, and MH-53 operational aircraft noise spectra were chosen for this study because they are representative of aircraft which are currently open to female aircrews and potentially vulnerable environments for female speech. The four noise conditions studied are representative of the typical range of noise spectra found at the pilot-copilot positions of the selected aircraft. Specifically, the four operational noise levels chosen for each aircraft consisted of an ambient noise condition of 66 dB and aircraft noise presented at 95 dB, 105 dB, and 115 dB.

During data collection, each member of the ten-subject listening panel was seated at an experimental test station, and one talker was seated at the talker test station in the VOCRES facility (Phases I and II) or the PACRAT facility (Phase III). The talker and the listeners were in the same noise environment during each experimental run. Each subject was equipped with the custom-fit headset or helmet corresponding to the experimental condition being evaluated (Table 1). For each experimental run, the word list appeared on the LED display in front of the talker, one word at a time (Figure 4a). The talker read each word, after which each member of the listening panel selected the word she/he believed was spoken from the list of six rhyming words

on their LED display (Figure 4b) by pressing the response button adjacent to that word. Data from each of the ten stations were sent simultaneously to a computer which calculated each listener's score for a specific talker for each experimental run. Data collection for Phases I-III of the study followed this procedure for each experimental run in all noise spectra and levels investigated. The study was conducted in a series of four phases in which specific variables were investigated at each phase.

**Phase IV**

All data were collected with the talker in a noise environment. As previously noted, the experimental design required the measurement of speech recognition performance using twenty talkers. Twenty talkers (usually only ten talkers are required) were selected to expand the applicability of the data and findings of the study. No listeners were needed for this phase of the study because the automatic speech recognition (ASR) systems "acted as the listeners."

The C-130E and MH-53 operational aircraft noise spectra were chosen for this study because they are currently open to female aircrews and potentially vulnerable for female speech. The two noise conditions studied are representative of the typical range of noise spectra found at the pilot-copilot positions of the selected aircraft. The two operational noise levels and spectra are the same as in the previous three phases of the study.

During testing of the ITT system, each talker was seated in a small sound booth (Figure 5) and high quality digital recordings were made to ensure that each ASR system would have the same input. These recordings were then played back to the IBM system. Each subject was equipped with a custom-fit headset corresponding to the experimental condition being evaluated (Table 1). The M-162 microphone was used in place of the M-87 microphone due to the findings in Phase II of this study. Each ASR system required a noise calibration procedure to set thresholds for speech detection. This consisted of the talker speaking one or more sentences before beginning each training or testing session at each noise level. For the ITT system each talker went through an enrollment phase to train the recognition system to recognize the words in the vocabulary. The IBM system did not require this enrollment phase since it is a speaker-independent system.

For each experimental run with the ITT system, a set of sentences appeared on the CRT in front of the talker. The talker read each sentence, and the computer recorded the output of the ASR. After each experimental run, the recognition accuracies were calculated by the computer. For each experimental run with the IBM system, the digital recordings were played back to the system instead of using live talkers. These procedures were followed for each experimental run in all noise spectra and levels investigated in Phase IV.

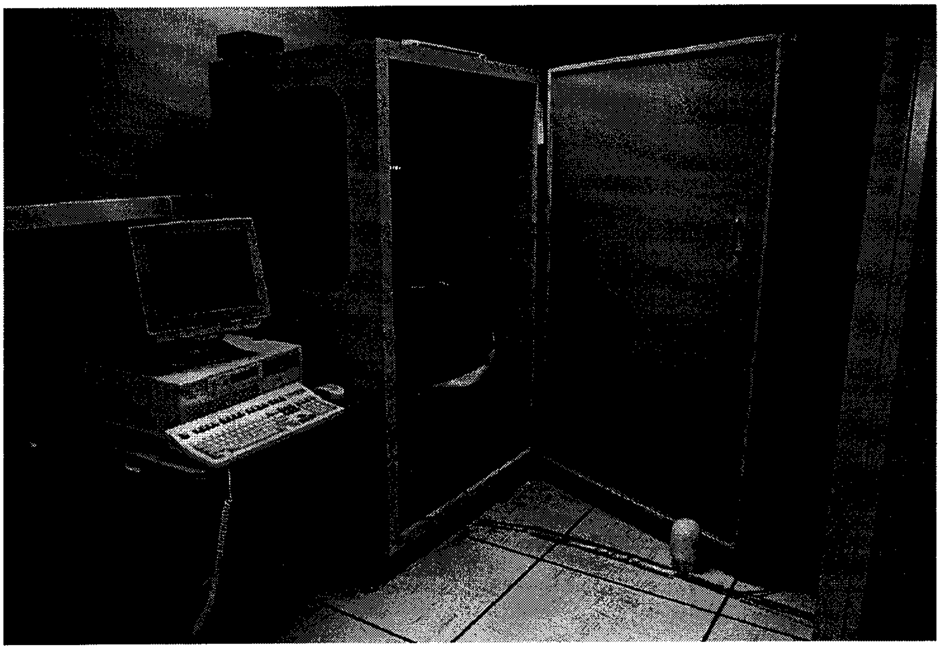


*Figure 5: Phase IV test setup with talker in sound booth.*

## EXPERIMENTAL PHASES

### Phase I

Phase I examined the influence of the spectrum and the level of four aircraft cockpit noises on the intelligibility of female and male speech. The three independent variables of subject, spectrum of noise, and level of noise were randomized to minimize effects such as variations in the repeat trials, subject differences, and learning. The dependent variable was percent correct speech intelligibility on the MRT. The operational noise spectra and levels previously noted were selected to identify potential areas requiring enhancements of female produced speech perceived by others in various noise environments. Phase I data were collected under a total of 320 conditions: four spectra × four levels for each spectrum × twenty talkers.

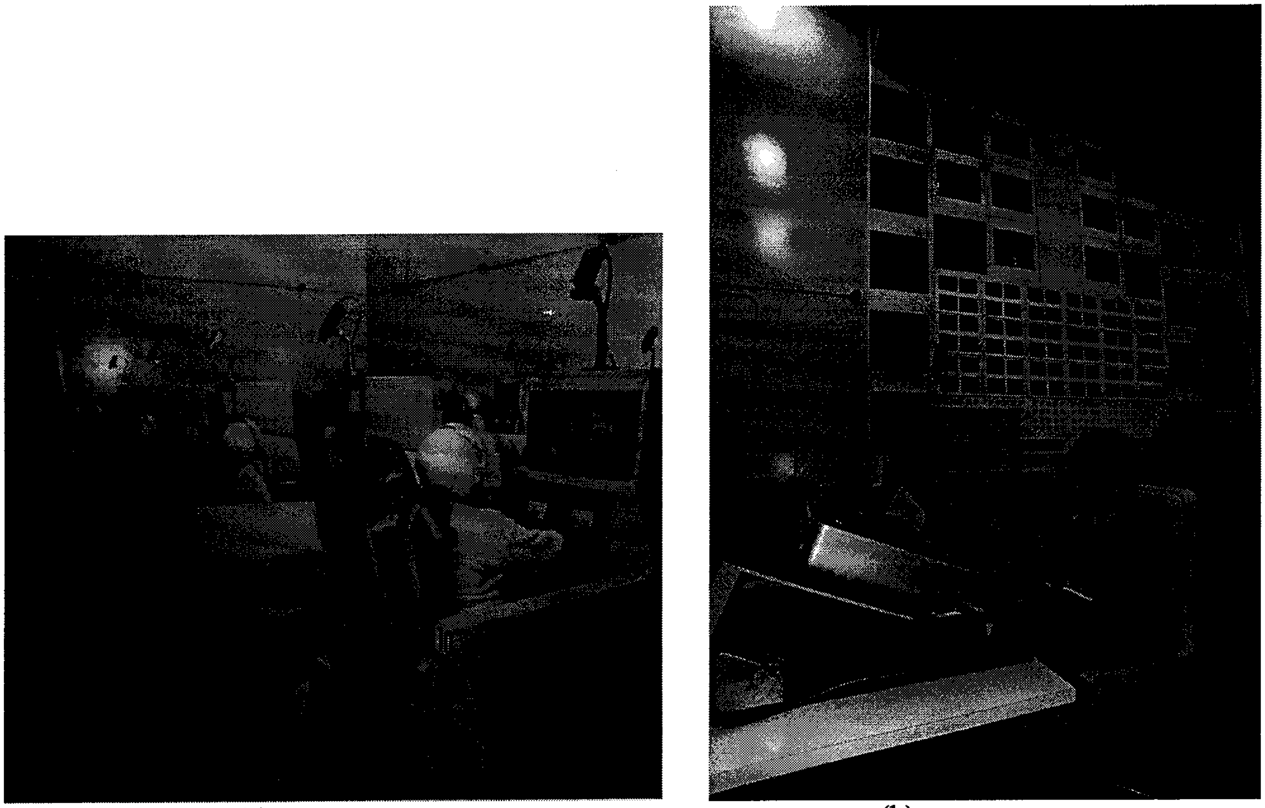
### Phase II

The purpose of Phase II was to identify the influence, if any, of noise-cancelling microphones on speech intelligibility. The independent variables investigated in Phase II were noise-cancelling microphones, noise spectrum, and noise level; the dependent variable was speech intelligibility. Two standard noise-cancelling microphones used for this phase were the M-87 boom microphone and the M-162 microphone. Intelligibility of male and female produced speech was measured using the M-162 microphone in three noise spectra: C-130E, C-141B, and MH-53, each at four noise levels. Data collected on the M-87 microphone in Phase I were extracted for

analysis in Phase II. The F-15A spectrum was not used in this phase because it requires the use of a helmet with an oxygen mask. The M-169 noise-cancelling microphone contained in the oxygen mask is the only microphone appropriate to use with the oxygen mask. Therefore, experimentation in Phase II included 240 total conditions: three spectra × four levels for each spectrum × twenty talkers for one microphone (M-87 microphone data were collected in Phase I).

**Phase III**

Phase III was conducted to evaluate the effect of digital coding of speech signals on speech intelligibility. Speech signals were encoded and decoded using the DoD standard LPC-10 and the CVSD vocoders. Speech intelligibility performance was evaluated with each system in the four noise spectra and four noise levels previously discussed. For this phase, the configuration of the experimental stations was varied slightly to better emulate the operational environment in which digital coding devices are used. The remote talker station was placed in the Performance and Communication Research and Technology (PACRAT) facility, a facility capable of generating noise spectra and levels identical to those used in VOCRES. PACRAT contains all of the features of VOCRES, plus task loading features. The ten stations in PACRAT are emulations of fighter aircraft cockpits and employ simultaneous dynamic performance tasks to load and overload the speech signals to determine their robustness (Figure 6). The talker in each experimental run was seated in the PACRAT facility in the same noise spectrum and level as the ten listeners seated in the VOCRES facility. The digitized speech signal was transmitted from the remote talker over phone lines via modem, coded and decoded by the system being used, and then received by the listeners who responded in the same manner as in all previous experimental phases. Phase III included a total of 640 experimental conditions: four spectra × four levels × two coder/vocoders × twenty talkers.



(a) (b)

*Figure 6: (a) PACRAT individual stations in the foreground. (b) PACRAT remote talker station.*

**Phase IV**

The purpose of Phase IV of this study was to measure the recognition accuracy of female and male speech using two state-of-the-art automatic speech recognition (ASR) systems in two noise spectra (C-130E and MH-53) in each of the four levels of noise used in previous phases. The two fundamentally different systems that were used were the ITT VRS-1290 and the IBM VoiceType ASR systems. Characteristics of these systems are described in the Facilities and Equipment section. The vocabulary used with these systems was developed during a joint Air Force-NASA in-flight study of voice control in the OV-10 aircraft (Appendix D). Subjects wore the H-157 headset with the M-162 microphone for this phase of the study. As in previous phases, twenty subjects were used as talkers; however, instead of the ten-subject listening panel the two ASR systems were used as the listeners. A total of 320 conditions were investigated in this phase: two spectra × four levels of each spectrum × two ASR systems × twenty talkers.

## RESULTS

Various measurement data are provided for each phase of the study. In Phase I and II, data are comprised of measurements of speech intelligibility of ten male and ten female talkers as perceived by a panel of ten listeners (five male and five female). In Phase III, data consist of

measurements of the intelligibility of coded and decoded speech of all talkers perceived by the panel of listeners. In Phase IV, data consist of the word and sentence recognition accuracy of two speech recognition systems. The responses of the individual subjects were averaged for each experimental condition. Means and standard deviations were calculated and differences among the means were evaluated using standard statistical paired t-tests at the 0.05 level.

In the following results, average percent correct intelligibility (the criterion measure) of the female speech is below that of the male speech in almost all conditions. These differences are relatively small, range from about one percent to ten percent, and may or may not be statistically significant. Further evaluations of these differences are discussed in later sections of this report.

Data were treated by measures of central tendency and variance with emphasis on the average differences between the means of the samples. The statistical significance of the differences between the means of the matched pairs (female and male) was determined by calculating the t-score and comparing it with the criterion t-value corresponding to the 95 percent confidence level (4). The calculated t-scores for each pair indicate the number of standard deviations separating the two means. If the t-score is greater than the criterion t-value at the 95 percent confidence level, the difference between the paired means is statistically significant. However, the statistically significant differences in many situations are so small they are indistinguishable in the operational situation. The critical issue is whether the performance in a particular condition is acceptable, marginal, or unacceptable. These performance levels are indicated by dashed horizontal lines across figures where applicable.

**Phase I**

*Aircraft Cockpit Noise Spectra*

The average intelligibility scores are summarized for the female and male subjects for each aircraft at the four levels of noise. The data are shown in graphical form in Figures 7 through 10 and in tabular form in Tables 2 through 5. The vertical bars on the figures represent plus and minus one standard deviation. Those differences between means that are statistically significant at the 95 percent level of confidence are circled on the graphs and are boxed in the tables. Dashed horizontal lines across figures indicate levels of acceptable, marginal, and unacceptable performance.

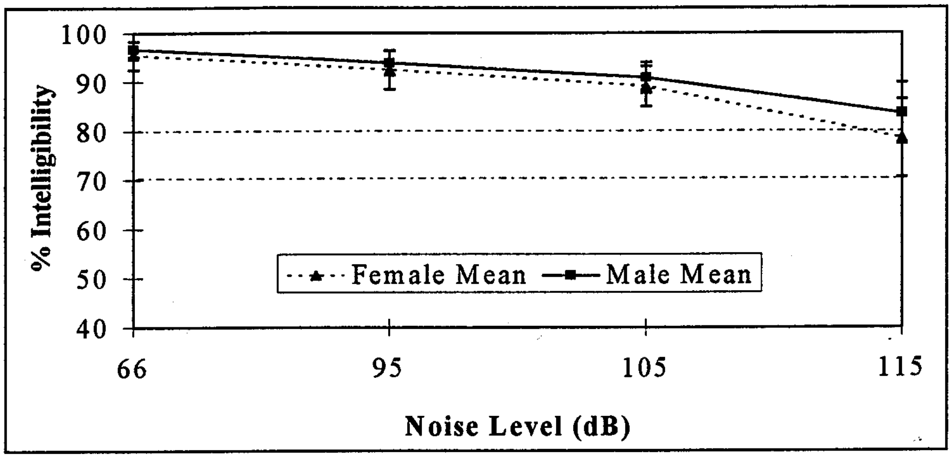



*Figure 7: Phase I - Male versus female intelligibility using C-130E spectrum, H-157 headset, and the M-87 microphone.*

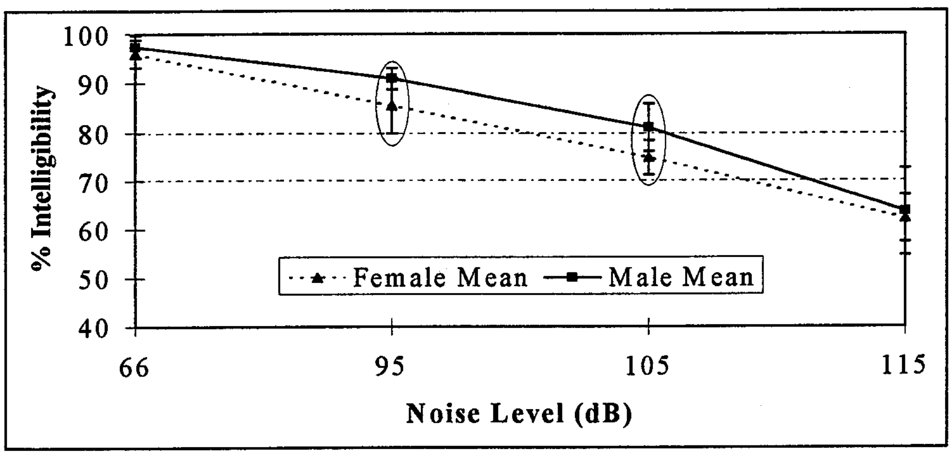



*Figure 8: Phase I - Male versus female intelligibility using C-141B spectrum, H-157 headset, and the M-87 microphone.*

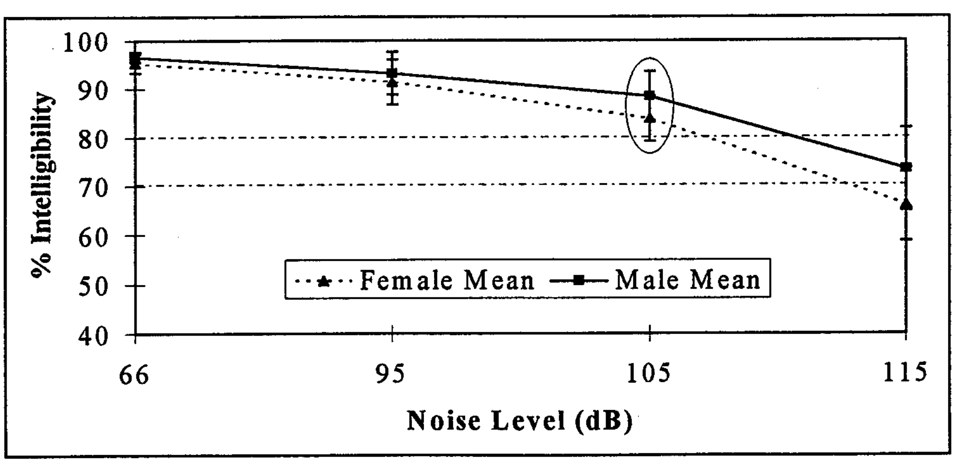



*Figure 9: Phase I - Male versus female intelligibility using F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and the M-169 microphone.*

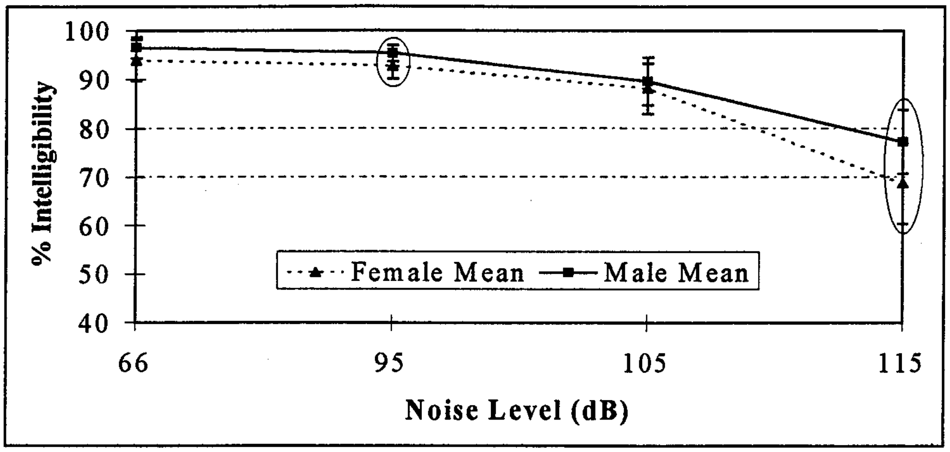



*Figure 10: Phase I - Male versus female intelligibility using MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 95.4 ± 2.4 | 92.4 ± 4.0 | 89.0 ± 4.1 | 78.5 ± 7.9 |
| Male - % avg. intelligibility ± standard deviation | 96.7 ± 1.6 | 93.8 ± 2.5 | 90.8 ± 3.1 | 83.6 ± 6.2 |
| Difference in Means | -1.3 | -1.4 | -1.8 | -5.1 |
| T-score | -1.29 | -0.95 | -1.1 | -1.6 |

*Table 2: Phase I - Male versus female intelligibility with C-130E spectrum, H-157 headset, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 96.0 ± 2.8 | 85.5 ± 5.6 | 74.9 ± 3.5 | 62.2 ± 4.8 |
| Male - % avg. intelligibility ± standard deviation | 97.5 ± 2.27 | 91.0 ± 2.1 | 81.1 ± 4.9 | 63.7 ± 9.0 |
| Difference in Means | -1.5 | -5.5 | -6.2 | -1.5 |
| T-score | -1.29 | -2.85 | -3.26 | -0.45 |

*Table 3: Phase I - Male versus female intelligibility with C-141B spectrum, H-157 headset, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 95.3 ± 2.0 | 91.4 ± 4.7 | 83.9 ± 4.7 | 66.1 ± 7.5 |
| Male - % avg. intelligibility ± standard deviation | 96.6 ± 1.1 | 93.2 ± 4.4 | 88.5 ± 5.1 | 73.4 ± 8.5 |
| Difference in Means | -1.3 | -1.8 | -4.6 | -7.3 |
| T-score | -1.89 | -0.92 | $\boxed{-2.10}$ | -2.02 |

*Table 4: Phase I - Male versus female intelligibility with F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and the M-169 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 93.9 ± 4.1 | 92.8 ± 2.6 | 88.2 ± 5.2 | 68.9 ± 8.3 |
| Male - % avg. intelligibility ± standard deviation | 96.5 ± 2.1 | 95.3 ± 1.6 | 89.7 ± 4.8 | 77.3 ± 6.5 |
| Difference in Means | -2.6 | -2.5 | -1.5 | -8.4 |
| T-score | -1.76 | $\boxed{-2.62}$ | -0.66 | $\boxed{-2.53}$ |

*Table 5: Phase I - Male versus female intelligibility with MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.*

Female speech perception was quantified in the cockpit noise environments of four aircraft in which female aviators are found. A matched experimental design was not implemented because three different headset/helmet communication systems were worn by the subjects in the four aircraft noise spectra. The effectiveness of the personal equipment systems interacts with the noise spectra and levels to influence the speech intelligibility. These interactions were not examined in this study.

The frequency bandwidth of standard Air Force voice communication systems is confined to approximately 300 Hz to 3500 Hz. Noise spectra with substantial energy in this speech frequency region, and slightly below, are most effective in masking the speech signal. In the ambient noise spectra at 66 dB, the speech signal was not masked and the intelligibility is essentially the same for all ambient conditions. The individual aircraft spectra were not presented in the ambient conditions; however, the subjects did wear the personal equipment items utilized during measurements in the four aircraft. Average intelligibility of male speech is 97 to 98 percent correct and of female speech is 94 to 96 percent. One hundred percent average intelligibility was not achieved, even under these ideal conditions.

The aircraft cockpit noise data in Figure 1 represent in-flight cruise conditions for which the spectra and level differ substantially among aircraft. In this study, the experimental conditions

presented all the spectra at the same four fixed overall sound pressure levels (OASPL). This was done to include the range of levels found in almost all operational aircraft, to allow comparisons among aircraft types, as well as to measure reductions in speech performance as levels of noise spectra were increased for the individual aircraft.

The relative influence of the different spectra on speech performance can be directly compared for the C-130E and C-141B conditions because experimental subjects wore the same headset-microphone communications equipment in both sets of measurements. The only difference between the experimental conditions was noise spectrum. The comparison is also of interest because the speech performance of both males and females is best in the C-130E and poorest in the C-141B. Speech performance is acceptable in all measured conditions for the C-130E and is unacceptable for both male and female speech at the highest level of the C-141B noise.

The noise spectrum of the C-141B is very flat with a slight rolloff starting at about 4000 Hz. The C-130E spectrum has a high peak around 63 Hz that is more than 15 dB greater than the next highest octave band level in the spectrum. The C-130E spectrum rolls off at about 5 dB per octave starting around 1000 Hz in the central region of the passband of the voice communication equipment. The C-130E overall level is determined by the peak level of 111 dB; the levels of the other octave bands are so far below the peak that they make no contribution to the overall level. When the two spectra are at the same OASPL, the C-141B spectrum is higher than the C-130E spectrum in all bands except 63 Hz where it is less. The C-130E is the less effective masker of the two because of the lower levels in almost all bands and the rolloff starting at 1000 Hz.

The decreases in intelligibility due to increases in level of the noise vary with aircraft spectrum. Also, the amount of the decrease becomes progressively larger with increasing levels of noise. For the C-130E, the decrease in intelligibility is three percent less at 105 dB than at 95 dB, and seven to 10 percent less at 115 than at 105 dB (Figure 7, Table 2). The C-141B intelligibility is 10 to 11 percent less at 105 dB than at 95 dB and 13 to 17 percent lower at 115 dB than at 105 dB (Figure 8, Table 3). These decreases in intelligibility are approximately the same for male and female speech except at the 115 dB, MH-53 condition where the decrease for females is larger and the 115 dB, C-141B where it is smaller.

*C-130E Aircraft*

Perception of the male and female speech is essentially the same at the 105 dB level of the C-130E noise and below with only a 5 percent difference at 115 dB (Figure 7, Table 2). None of the differences are statistically significant. Both male and female speech are around the 90 percent correct region and above at noise levels of 105 dB and below. At 115 dB, the accuracies are 79 percent correct for females and 84 percent correct for the males; both are acceptable. The overall level of the noise of the C-130E during maximum endurance cruise is about 111 dB in the flight crew compartment and a maximum level of 115 dB at one of the other crew stations (5). Voice communication conditions in this aircraft, for female and male talkers, are considered acceptable.

*C-141B Aircraft*

The speech intelligibility of both males and females is vulnerable to this noise spectrum, dropping in mean intelligibility almost 40 percent from the ambient to the 115 dB noise condition (Figure 8, Table 3). The mean differences between genders at both the 95 dB and 105 dB noise conditions are statistically significant at the 95 percent confidence level. Both female and male speech are acceptable at the 95 dB level; at 105 dB, male speech is acceptable and female speech is marginal; and both are unacceptable at the 115 dB level. Assuming that the relatively linear function shown by the graph is reliable, the extrapolated percent correct intelligibility at 100 dB should be almost 80 percent for the female and acceptable; it should be higher at lower levels of noise. The overall level of the noise measured between the pilot and copilot on the C-141B is almost 96 dB during cruise with a worst-case condition of 117 dB during taxi with four engines at taxi power and 111 dB during climb to 3000 feet (6). Therefore, communication will be acceptable during cruise, but unacceptable during taxi and climb.

*F-15A Aircraft*

Speech perception decreases and the differences between female and male mean speech intelligibility increase in the F-15A as the level of the noise increases (Figure 9, Table 4). The only statistically significant difference between the mean values occurs at the 105 dB noise condition which is acceptable for both genders. At the 115 dB level of noise, the male speech is marginal and the female speech unacceptable. The overall sound pressure level of the F-15A cockpit noise during cruise is about 110 dB and during a high speed run it is about 115 dB (7). The data suggest that female speech perception is marginal to unacceptable in the high noise environments of these two flight conditions and that the male operates in the marginal region at 115 dB. It is presumed that experienced aviators compensate to maintain communications for marginal situations when the maximum levels of noise are encountered. However, improvement is required for female speech to be understood by other aviators in the 110 dB - 115 dB levels of noise.

*MH-53 Helicopter*

The mean intelligibility response curves are similar for the MH-53 helicopter (Figure 10) and the F-15A fighter aircraft (Figure 9) with the scores in the helicopter noise slightly better (10). Statistically significant differences between male and female speech perception occur at the 95 dB and 115 dB noise conditions. The small difference of only about 2.5 percent at the 95 dB noise condition is statistically significant because the standard deviations are very small; therefore, this difference is not operationally significant. The mean difference at the 115 dB level of noise is about 8 percent. The noise spectra of the MH-53 and the C-130E vehicles are very similar except for the peak at 63 Hz in the C-130E spectrum. The maximum level of the noise measured between the pilot and copilot during cruise is 111 dB, while under maximum cruise it is 115 dB. The speech perception of both female and male is acceptable at all except the 115 dB condition. At 115 dB, male speech is in the marginal region, close to the acceptable range. The female speech is a little below the marginal region and must continue to be considered unacceptable.

Improvement in female speech perception is required in these high level noise environments for good recognition by other aviators.

## Phase II

### *Noise-Cancelling Microphones*

The conditions in Phase I in which the M-87 noise-cancelling microphone was used were repeated in Phase II with the M-162 noise-cancelling microphone. These two sets of data (Phase I M-87 microphone and Phase II M-162 microphone) were compared to evaluate the relative effectiveness in noise of the microphones with female and male produced speech. The two independent variables of aircraft noise spectrum and level were randomized to minimize effects due to uncontrolled variance. The dependent variable was percent correct speech intelligibility on the MRT. The M-169 oxygen mask noise-cancelling microphone was not included in this evaluation. Since no alternative mask microphone is available, the M-169 data collected in Phase I represent its performance in the spectra and levels of the noises of interest.

The average speech intelligibility for the M-162 microphone in the various levels of the aircraft spectra are shown in graphical form in Figures 11 through 13 and in tabular form in Tables 6 through 8. No statistically significant differences between female and male speech were observed with the M-162 microphone. All performance was acceptable, according to the performance criteria, except for the 115 dB noise condition for the C-141B aircraft which was unacceptable for both male and female talkers.

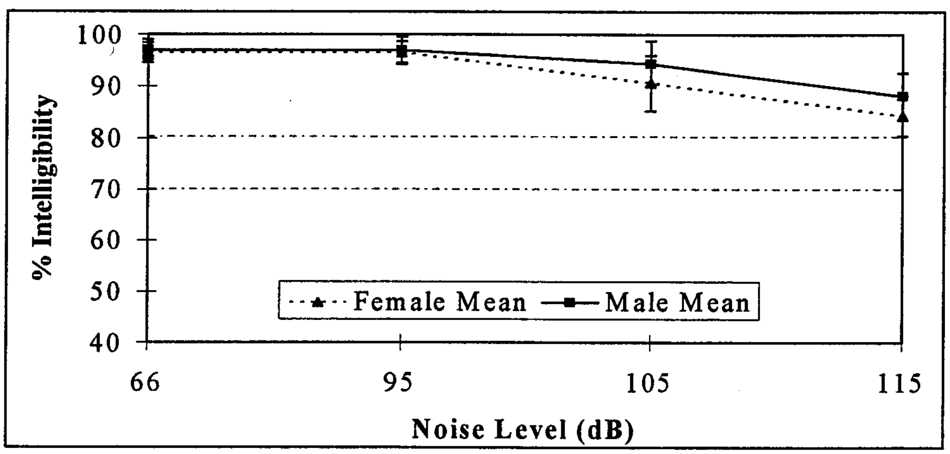



*Figure 11: Phase II - Male versus female intelligibility with C-130E spectrum, H-157 headset, and the M-162 microphone.*

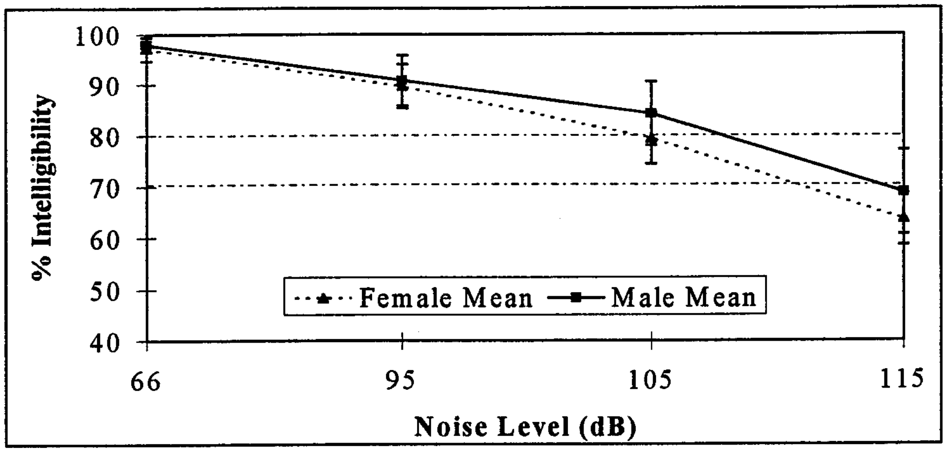



*Figure 12: Phase II - Male versus female intelligibility with C-141B spectrum, H-157 headset, and the M-162 microphone.*

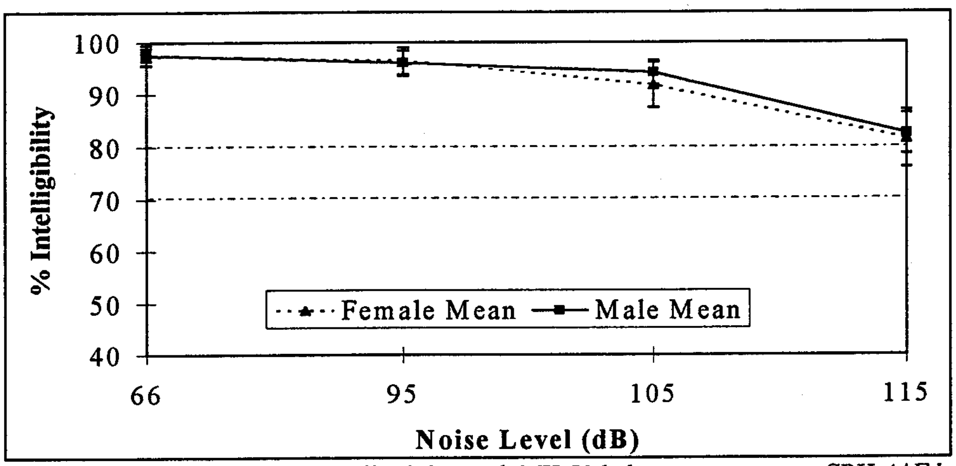



*Figure 13: Phase II - Male versus female intelligibility with MH-53 helicopter spectrum, SPH-4AF helmet, and the M-162 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 96.5 ± 1.9 | 96.5 ± 2.2 | 90.6 ± 5.3 | 84.3 ± 3.9 |
| Male - % avg. intelligibility ± standard deviation | 97.1 ± 1.9 | 97.0 ± 2.6 | 94.4 ± 4.4 | 88.1 ± 4.6 |
| Difference in Means | -0.6 | -0.5 | -3.8 | -3.8 |
| T-score | -0.73 | -0.53 | -1.73 | -1.97 |

*Table 6: Phase II - Male versus female intelligibility with C-130E spectrum, H-157 headset, and the M-162 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 97.1 ± 2.4 | 89.7 ± 4.4 | 79.4 ± 5.1 | 63.6 ± 5.1 |
| Male - % avg. intelligibility ± standard deviation | 97.9 ±1.4 | 90.8 ± 5.0 | 84.3 ± 6.4 | 68.9 ± 8.3 |
| Difference in Means | -0.8 | -1.1 | -4.9 | -5.3 |
| T-score | -0.94 | -0.52 | -1.89 | -1.7 |

*Table 7: Phase II - Male versus female intelligibility with C-141B spectrum, H-157 headset, and the M-162 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 97.2 ± 1.7 | 96.5 ± 2.6 | 91.9 ± 4.3 | 81.6 ± 5.5 |
| Male - % avg. intelligibility ± standard deviation | 97.6 ± 1.9 | 96.1 ± 2.4 | 94.3 ± 2.3 | 82.6 ± 3.8 |
| Difference in Means | -0.4 | 0.4 | -2.4 | -1.0 |
| T-score | -0.54 | 0.35 | -1.56 | -0.45 |

*Table 8: Phase II - Male versus female intelligibility with MH-53 helicopter spectrum, SPH-4AF helmet, and the M-162 microphone.*

Performance of the M-87 and the M-162 microphones with female produced speech is shown in Figures 14 through 16 and Tables 9 through 11 and for male speech in Figures 17 through 19 and Tables 12 through 14. Mean speech intelligibility with the M-162 is better than with the M-87 for all aircraft and all levels of noise. Female and male speech perception with the M-162 is acceptable in all conditions except the C-141B at the 115 dB level of noise. Female speech performance with the M-87 is marginal, and with the M-162 is acceptable in the C-130E at the 115 dB level of noise. Both microphones are unacceptable for the C-141B 115 dB noise condition and the M-87 is marginal in the MH-53 helicopter spectrum at 115 dB of noise.

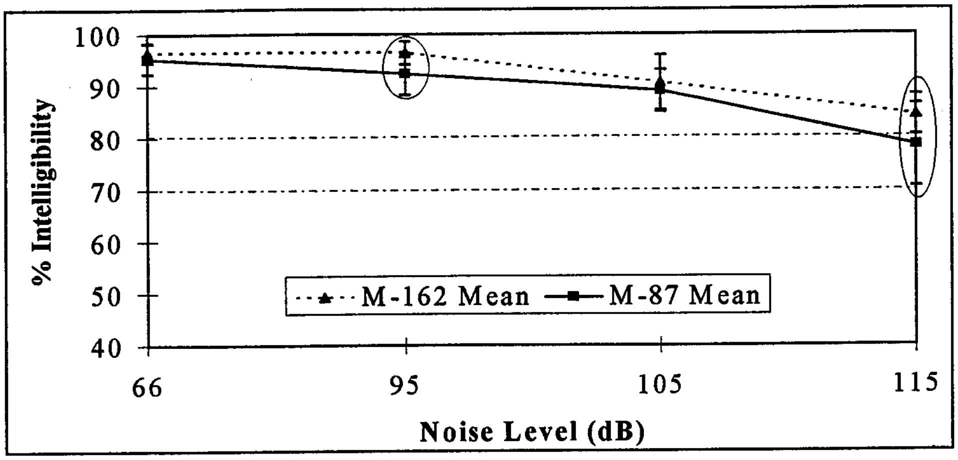



*Figure 14: Phase II - M-162 versus M-87 microphone with the C-130E spectrum, H-157 headset, and female subjects.*

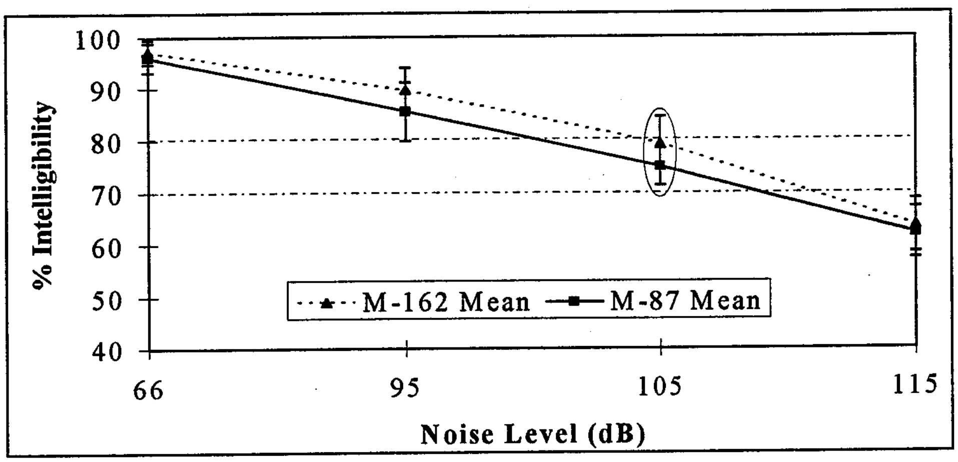



*Figure 15: Phase II - M-162 versus M-87 microphone with the C-141B spectrum, H-157 headset, and female subjects.*

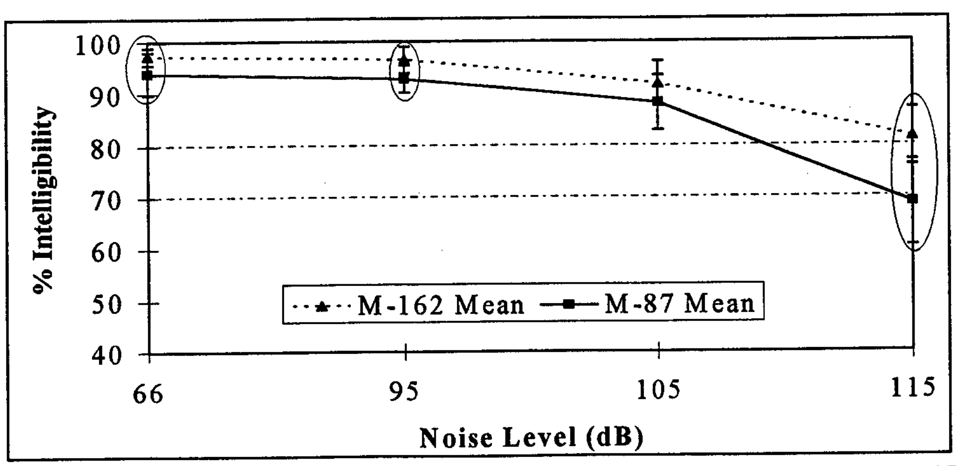



*Figure 16: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and female subjects.*

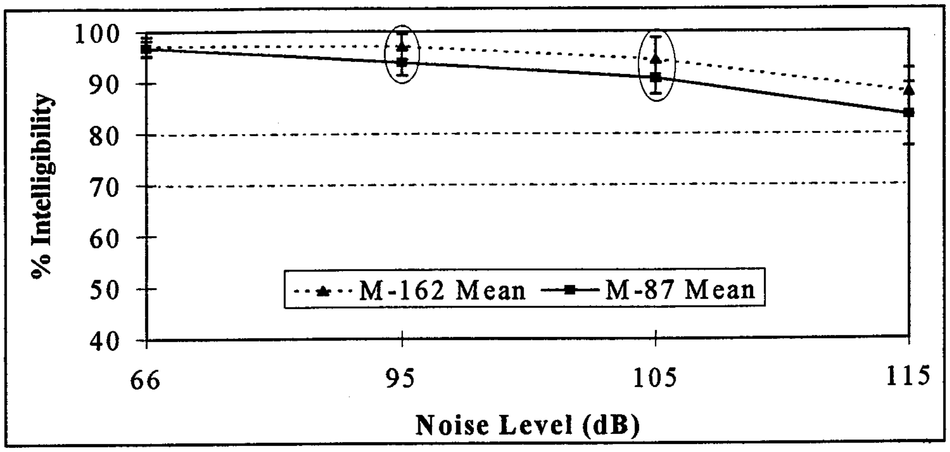



*Figure 17: Phase II - M-162 versus M-87 microphone with the C-130E spectrum, H-157 headset, and male subjects.*

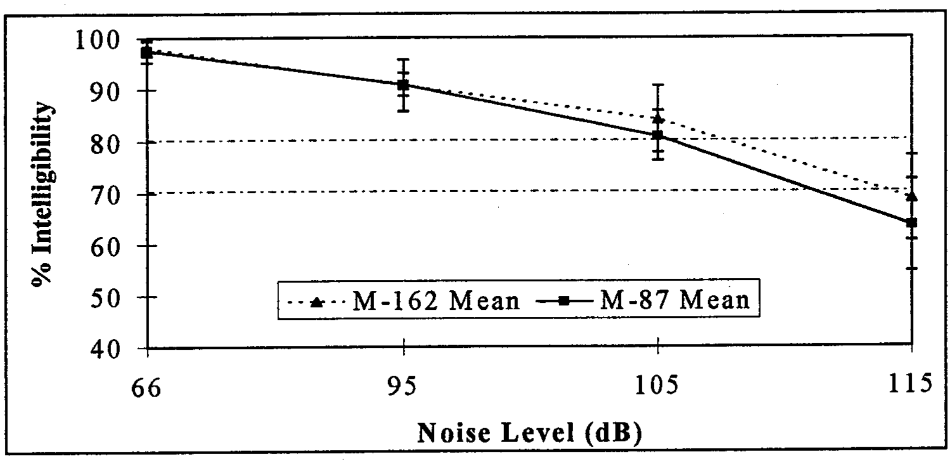



*Figure 18: Phase II - M-162 versus M-87 microphone with the C-141B spectrum, H-157 headset, and male subjects.*

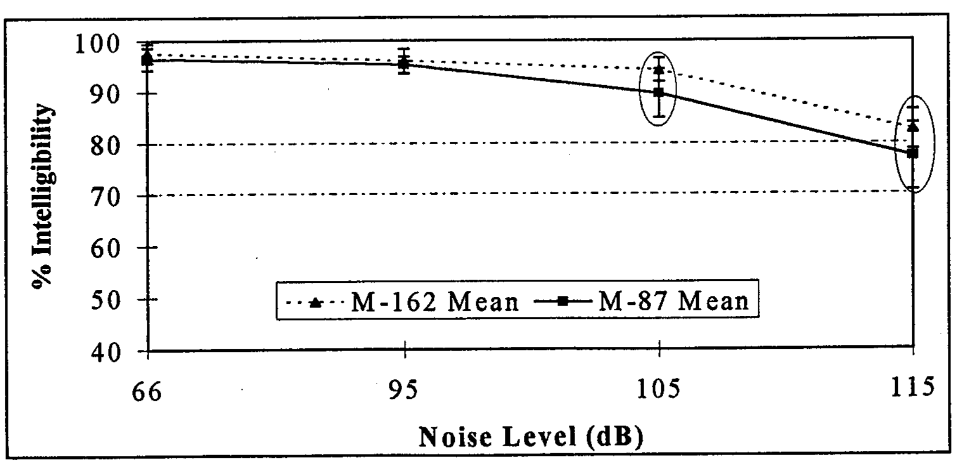



*Figure 19: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and male subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| M-162 - % avg. intelligibility ± standard deviation | 96.5 ± 1.9 | 96.5 ± 2.2 | 90.6 ± 5.3 | 84.3 ± 3.9 |
| M-87 - % avg. intelligibility ± standard deviation | 95.4 ± 2.9 | 92.4 ± 4.0 | 89.0 ± 4.1 | 78.5 ± 7.9 |
| Difference in Means | 1.1 | 4.1 | 1.6 | 5.8 |
| T-score | 1.00 | 2.78 | 0.77 | 2.11 |

*Table 9: Phase II - M-162 versus M-87 microphone with the C-130E spectrum, H-157 headset, and female subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| M-162 - % avg. intelligibility ± standard deviation | 97.1 ± 2.4 | 89.7 ± 4.4 | 79.4 ± 5.1 | 63.6 ± 5.1 |
| M-87 - % avg. intelligibility ± standard deviation | 96.0 ± 2.8 | 85.5 ± 5.6 | 74.9 ± 3.5 | 62.2 ± 4.8 |
| Difference in Means | 1.1 | 4.2 | 4.5 | 1.4 |
| T-score | 0.97 | 1.85 | 2.32 | 0.63 |

*Table 10: Phase II - M-162 versus M-87 microphone with the C-141B spectrum, H-157 headset, and female subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| M-162 - % avg. intelligibility ± standard deviation | 97.2 ± 1.7 | 96.5 ± 2.6 | 91.9 ± 4.3 | 81.6 ± 5.5 |
| M-87 - % avg. intelligibility ± standard deviation | 93.9 ± 4.1 | 92.8 ± 2.6 | 88.2 ± 5.2 | 68.9 ± 8.3 |
| Difference in Means | 3.3 | 3.6 | 3.7 | 12.7 |
| T-score | 2.31 | 3.12 | 1.72 | 4.06 |

*Table 11: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and female subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| M-162 - % avg. intelligibility ± standard deviation | 97.1 ± 1.9 | 97.0 ± 2.6 | 94.4 ± 4.4 | 88.1 ± 4.6 |
| M-87 - % avg. intelligibility ± standard deviation | 96.7 ± 1.6 | 93.8 ± 2.5 | 90.8 ± 3.1 | 83.6 ± 6.2 |
| Difference in Means | 0.4 | 3.2 | 3.6 | 4.5 |
| T-score | 0.50 | 2.82 | 2.12 | 1.85 |

*Table 12: Phase II - M-162 versus M-87 microphone with the C-130E spectrum, H-157 headset, and male subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| M-162 - % avg. intelligibility ± standard deviation | 97.9 ± 1.4 | 90.8 ± 5.0 | 84.3 ± 6.4 | 68.9 ± 8.3 |
| M-87 - % avg. intelligibility ± standard deviation | 97.5 ± 2.3 | 91.0 ± 2.1 | 81.1 ± 4.9 | 63.7 ± 9.0 |
| Difference in Means | 0.4 | -0.2 | 3.2 | 5.2 |
| T-score | 0.57 | -0.10 | 1.27 | 1.35 |

*Table 13: Phase II - M-162 versus M-87 microphone with the C-141B spectrum, H-157 headset, and male subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| M-162 - % avg. intelligibility ± standard deviation | 97.6 ± 1.9 | 96.1 ± 2.4 | 94.3 ± 2.3 | 82.6 ± 3.8 |
| M-87 - % avg. intelligibility ± standard deviation | 96.5 ± 2.1 | 95.3 ± 1.6 | 89.7 ± 4.8 | 77.3 ± 6.5 |
| Difference in Means | 1.1 | 0.8 | 4.6 | 5.3 |
| T-score | 1.24 | 0.79 | 2.72 | 2.19 |

*Table 14: Phase II - M-162 versus M-87 microphone with the MH-53 helicopter spectrum, SPH-4AF helmet, and male subjects.*

The microphone conditions for which the mean differences were statistically significant at the 95 percent confidence level exhibited no pattern relative to the noise conditions or subjects. The general patterns of percent correct intelligibility showed reduced intelligibility with increased level of noise. The performance of the M-162 microphone exceeded that of the M-87 microphone in all conditions by a margin of about 5 percent, except for the MH-53 noise condition of 115 dB where it was 12 percent. The t-scores for these conditions are relatively low

and close to the critical t-value, and the statistical significance is influenced by the variance of the data. The t-scores decrease as the variance increases for the same number of subjects.

The Phase II data indicate that the mean female speech perception is lower than the mean male speech perception for both microphones in all conditions; however, the amount of difference is relatively small and not statistically significant. As seen in Figure 20, the intelligibility of female speech is about 12 percent better with the M-162 microphone than with the M-87. This improvement in speech intelligibility measured with the M-162 is evident in all the experimental conditions for the C-130E, the C-141B, and the MH-53 helicopter (Figures 20-22). These data suggest that the perception of both female and male speech may be improved in the three aircraft spectra at all noise conditions by replacing the M-87 microphone with the M-162 microphone.

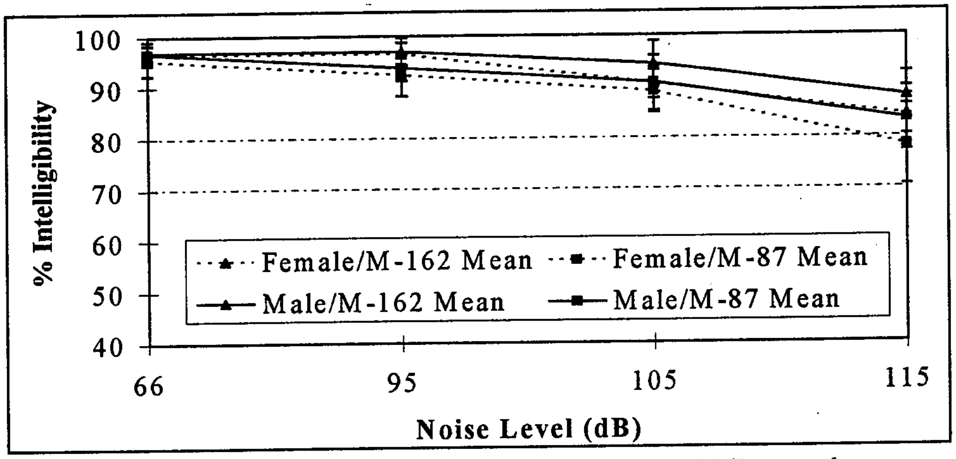



*Figure 20: Phase II - Male versus female with C-130E spectrum and M-87/M-162 microphones*

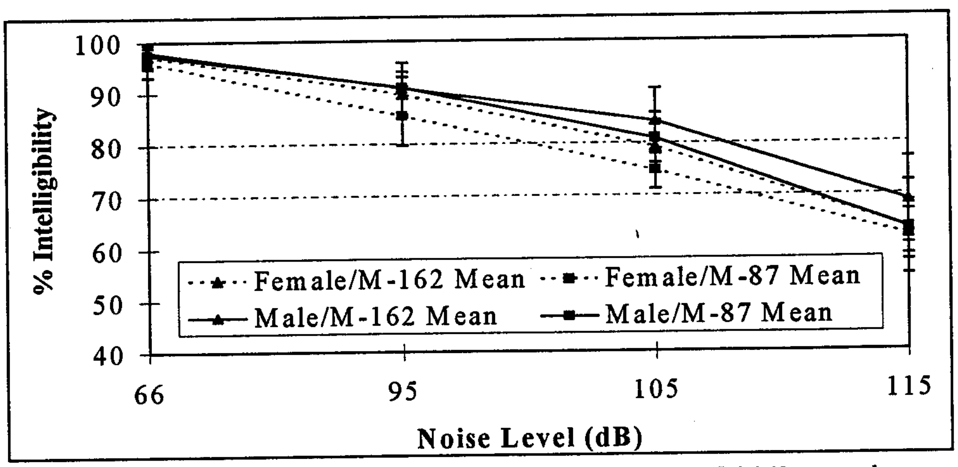



*Figure 21: Phase II - Male versus female with the C-141B spectrum and M-87/M-162 microphones*

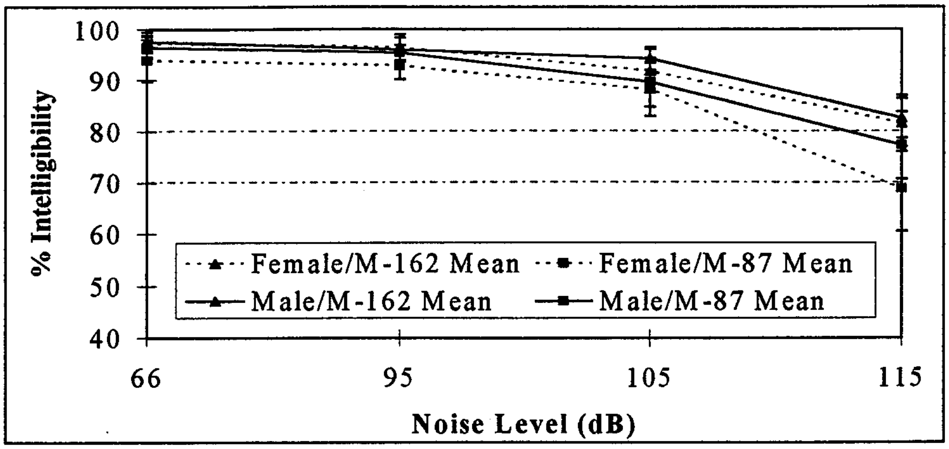



*Figure 22: Phase II - Male versus female with the MH-53 spectrum and the M-87/M-162 microphones.*

**Phase III**

*Speech Coders*

In Phase III, a remote talker station was located in the PACRAT voice communication research facility (Figure 6) that contains the same stimulus and response capabilities as in the VOCRES (Figure 2), including a programmable high intensity sound system. The talker was seated at this remote communication station in PACRAT (Figure 6) and the listeners remained at their individual stations in VOCRES. As in previous phases, all talkers and listeners wore the headset/helmet communications equipment appropriate for each noise condition (Table 1). The listeners at their individual stations responded by pressing subject response buttons corresponding to their perception of the received speech signals. Data were collected for vocoded female and male speech in four aircraft noise spectra at four levels of each of the noises.

*Linear Predictive Coder (LPC)*

Speech intelligibility data for the LPC-10 coded female and male speech in the various aircraft noise environments are summarized in Figures 23 through 26 and Tables 15 through 18. In the ambient and 95 dB noise levels, there were no statistically significant differences between the perception of the female and the male speech. All of the aircraft communications were acceptable in the ambient condition, ranging from 80 to 85 percent correct responses. All of the aircraft communications are marginal in the noises at the levels of 95 dB, ranging from 72 to 78 percent correct.

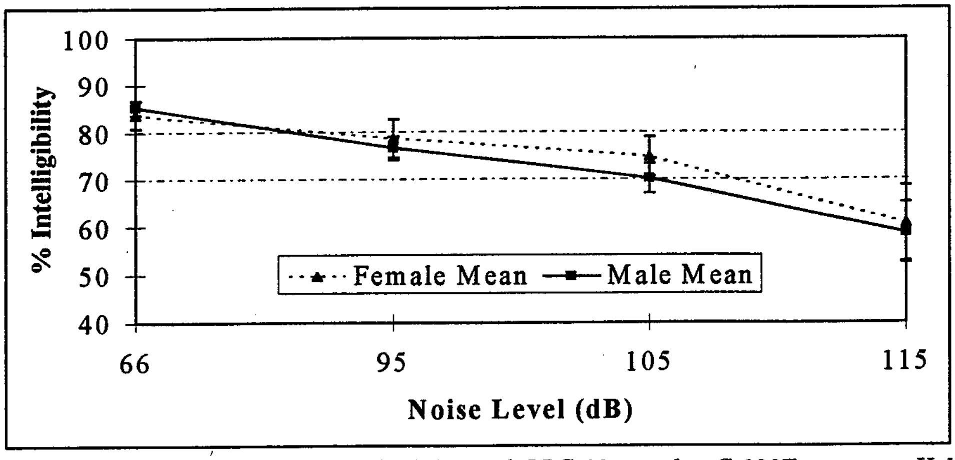



*Figure 23: Phase III - Male versus female intelligibility with LPC-10 vocoder, C-130E spectrum, H-157 headset, and the M-87 microphone.*

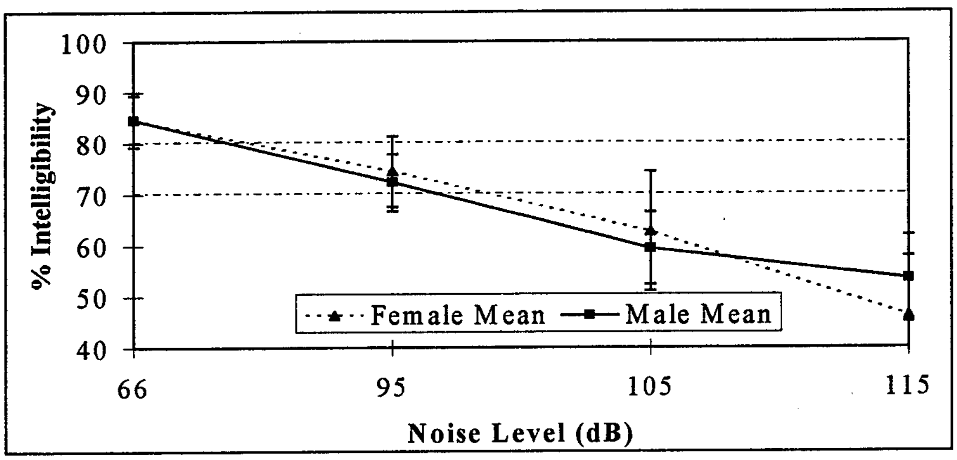



*Figure 24: Phase III - Male versus female intelligibility with LPC-10 vocoder, C-141B spectrum, H-157 headset, and the M-87 microphone.*

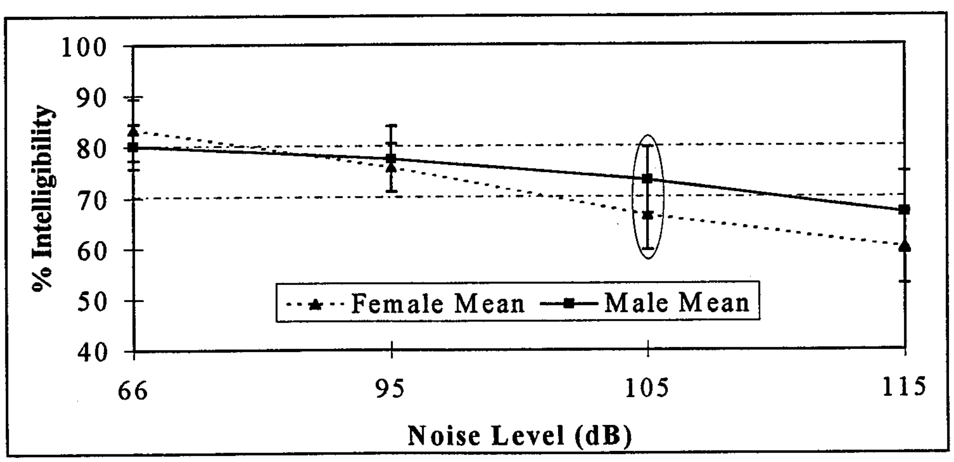



*Figure 25: Phase III - Male versus female intelligibility with LPC-10 vocoder, F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and the M-169 microphone.*

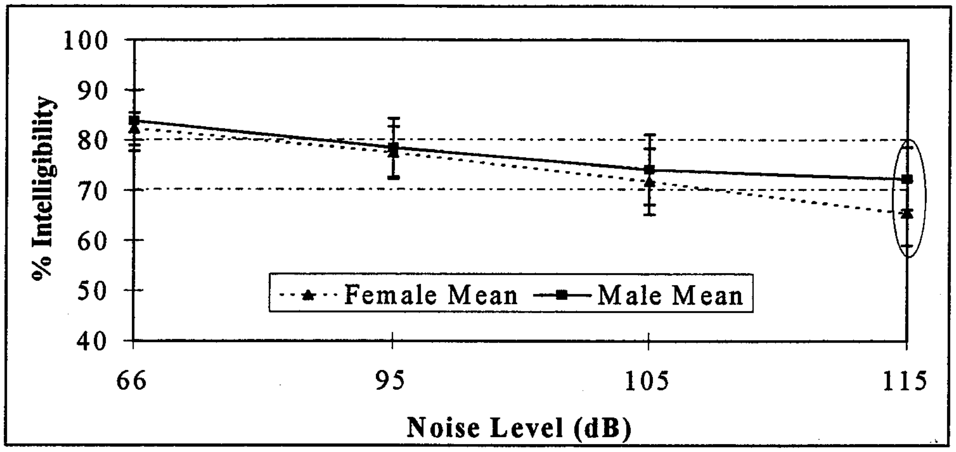



*Figure 26: Phase III - Male versus female intelligibility with LPC-10 vocoder, MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 83.7 ± 3.8 | 78.7± 4.5 | 74.8 ± 6.5 | 60.7 ± 9.7 |
| Male - % avg. intelligibility ± standard deviation | 85.3 ± 5.4 | 76.7 ± 4.4 | 70.2 ± 8.8 | 58.8 ± 11.3 |
| Difference in Means | -1.6 | 2.0 | 4.6 | 1.9 |
| T-score | -0.79 | 1.02 | 1.32 | 0.41 |

*Table 15: Phase III - Male versus female intelligibility with LPC-10 vocoder, C-130E spectrum, H-157 headset, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 84.6 ± 5.4 | 74.4 ± 7.0 | 62.8 ± 11.6 | 46.3 ± 11.6 |
| Male - % avg. intelligibility ± standard deviation | 84.7 ± 4.8 | 72.3 ± 5.6 | 59.4 ± 7.1 | 53.5 ± 8.4 |
| Difference in Means | -0.1 | 2.2 | 3.4 | -7.2 |
| T-score | -0.04 | 0.77 | 0.77 | -1.58 |

*Table 16: Phase III - Male versus female intelligibility with LPC-10 vocoder, C-141B spectrum, H-157 headset, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 83.3 ± 6.0 | 76.0 ± 4.7 | 66.5 ± 6.8 | 60.1 ± 7.1 |
| Male - % avg. intelligibility ± standard deviation | 80.1 ± 4.4 | 77.6 ± 6.42 | 73.5 ±6.5 | 67.1 ± 8.1 |
| Difference in Means | 3.2 | -1.6 | -7.0 | -7.0 |
| T-score | 1.38 | -0.67 | -2.34 | -2.05 |

*Table 17: Phase III - Male versus female intelligibility with LPC-10 vocoder, F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and the M-169 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 82.3 ± 3.2 | 77.5 ± 5.1 | 71.8 ± 6.6 | 65.4 ± 6.5 |
| Male - % avg. intelligibility ± standard deviation | 83.9 ± 6.0 | 78.5 ± 5.8 | 74.2 ±7.0 | 72.4 ± 6.2 |
| Difference in Means | -1.4 | -1.0 | -2.4 | -7.0 |
| T-score | -0.71 | -0.39 | -0.79 | -2.46 |

*Table 18: Phase III - Male versus female intelligibility with LPC-10 vocoder, MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.*

For the F-15A noise at a level of 105 dB, the female-male difference of about 6 percent is statistically significant, with better perception of the male speech. Communications are no better than marginal for all aircraft noise conditions at 105 dB, except the F-15A where female speech is unacceptable, and the C-141B where the speech of both genders is unacceptable. Average values range from 59 to 74 percent correct. Voice communications are unacceptable for all conditions in the noises at 115 dB, ranging from 46 to 67 percent correct, except for the male speech that is marginal in the MH-53 noise.

Overall, the perception of LPC-10 coded female and male speech is acceptable in the ambient condition, marginal in the 95 dB level, and unacceptable in the 105 and 115 dB levels of the noises. Perception of the female and male speech is very similar in the lower levels of the noise. At the higher levels of noise, the female speech tends to be a little less intelligible than the male speech; however, lesser intelligibility is statistically significant at only the two conditions cited earlier.

*Continuously Variable Delta Slope Modulation Coder (CVSD)*

Speech intelligibility data for the CVSD coded speech in the various aircraft noise environments are summarized in Figures 27 through 30 and Tables 19 through 22. For 66 and 95 dB, there are no statistically significant differences between the perception of the female and male speech. Generally, these voice communications are acceptable with values ranging from about 77 to 94 percent correct.

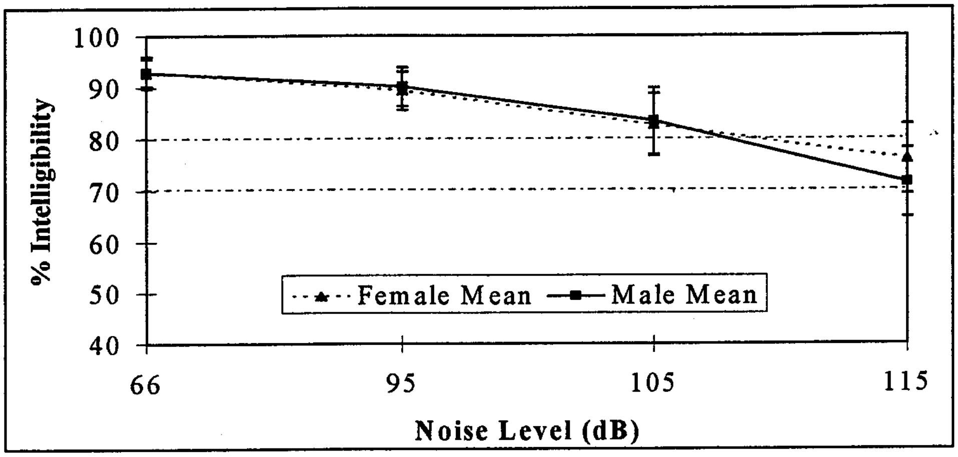



*Figure 27: Phase III - Male versus female intelligibility with CVSD vocoder, C-130E spectrum, H-157 headset, and the M-87 microphone.*

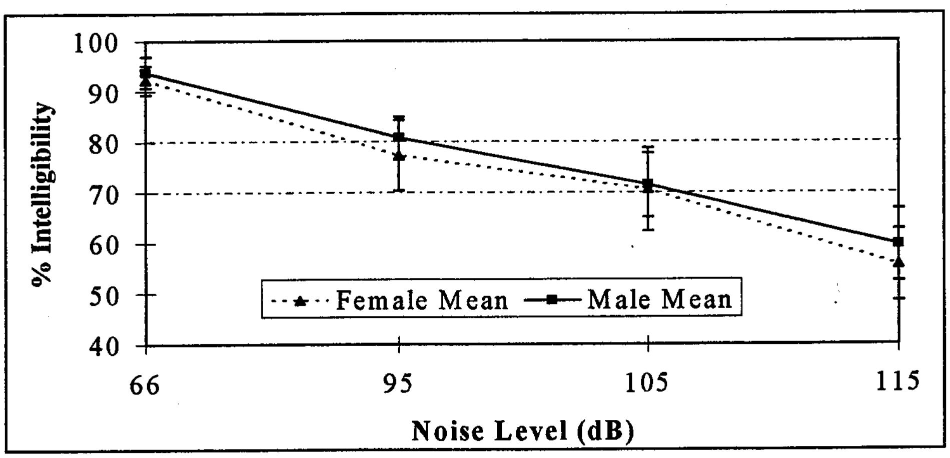



*Figure 28: Phase III - Male versus female intelligibility with CVSD vocoder, C-141B spectrum, H-157 headset, and the M-87 microphone.*

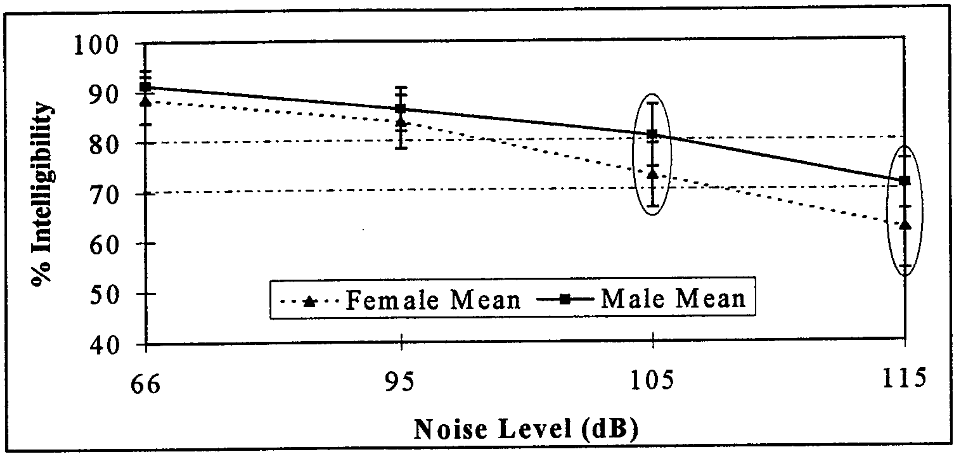



*Figure 29: Phase III - Male versus female intelligibility with CVSD vocoder, F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and the M-169 microphone.*

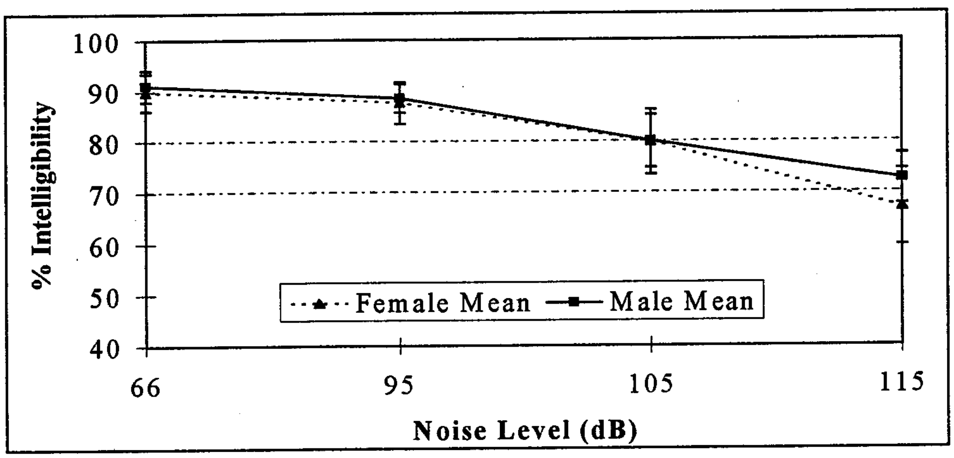



*Figure 30: Phase III - Male versus female intelligibility with CVSD vocoder, MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 92.8 ± 3.2 | 89.3 ± 3.7 | 82.7 ± 5.9 | 76.1 ± 6.8 |
| Male - % avg. intelligibility ± standard deviation | 92.9 ± 2.7 | 90.1 ± 3.7 | 83.4 ± 6.4 | 71.6 ± 6.7 |
| Difference in Means | -0.1 | -0.8 | -0.7 | -4.5 |
| T-score | -0.02 | -0.48 | -0.25 | 1.50 |

*Table 19: Phase III - Male versus female intelligibility with CVSD vocoder, C-130E spectrum, H-157 headset, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 92.3 ± 2.8 | 77.4 ± 6.9 | 70.6 ± 8.2 | 55.7 ± 7.1 |
| Male - % avg. intelligibility ± standard deviation | 93.8 ± 3.1 | 80.9 ± 4.2 | 71.5 ± 6.4 | 59.6 ± 7.2 |
| Difference in Means | -1.5 | -3.5 | -0.9 | -3.9 |
| T-score | -1.16 | -1.36 | -0.27 | -1.23 |

*Table 20: Phase III - Male versus female intelligibility with CVSD vocoder, C-141B spectrum, H-157 headset, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 88.5 ± 4.7 | 83.9 ± 5.3 | 73.1 ± 6.42 | 62.5 ± 8.2 |
| Male - % avg. intelligibility ± standard deviation | 91.2 ± 3.2 | 86.4 ± 4.3 | 81.0 ± 6.2 | 71.3 ± 5.0 |
| Difference in Means | -2.8 | -2.50 | -7.9 | -8.8 |
| T-score | -1.53 | -1.18 | -2.79 | -2.90 |

*Table 21: Phase III - Male versus female intelligibility with CVSD vocoder, F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and the M-169 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. intelligibility ± standard deviation | 89.8 ± 3.7 | 87.6 ± 4.1 | 79.9 ± 6.4 | 67.1 ± 7.5 |
| Male - % avg. intelligibility ± standard deviation | 91.0 ± 3.1 | 88.6 ± 2.7 | 80.0 ± 5.1 | 72.6 ± 5.0 |
| Difference in Means | -1.2 | -1.0 | -0.1 | -5.5 |
| T-score | -0.79 | -0.60 | -0.03 | -1.98 |

*Table 22: Phase III - Male versus female intelligibility with CVSD vocoder, MH-53 helicopter spectrum, SPH-4AF helmet, and the M-87 microphone.*

The only statistically significant differences between the female and male speech are with the F-15A noise at levels of 105 and 115 dB. At 105 dB, male speech is acceptable and female speech is marginal; at noise levels of 115 dB, male speech is marginal and female speech is unacceptable. The perception of female and male speech is virtually the same in the other three aircraft noises at all four levels. Both female and male speech are unacceptable for all aircraft noises at 115 dB, except for the C-130E where both are marginal. As with most other factors

examined in different levels of noise, perception of female speech tends to decrease more than male speech as the levels of the noises increase to the highest measured levels.

*LPC-10 and CVSD Performance*

The perception of the female speech is almost the same as the male speech when both are processed by either one or the other vocoder. Only four of the thirty-two measurement conditions show statistically significant differences in percent correct intelligibility due to gender. Two of these are with CVSD speech at the higher noise levels of the F-15A aircraft and two are with LPC-10 speech in 105 dB of F-15A noise and 115 dB of MH-53 aircraft noise. Overall, the perception of the female speech is equally as effective as the male speech for either vocoder.

Although the intelligibility of the female and male speech is very similar for either vocoder, the differences between the performance of the two vocoders are statistically significant at almost all conditions (Figures 31 through 38 and Tables 23 through 30). In all test conditions, the average percent correct intelligibility of the CVSD speech is higher than the LPC-10 data, revealing that the CVSD speech in quiet and noise is more intelligible than the LPC-10 speech. Differences between the vocoder mean values as a function of level of the noises are as high as 15 percent. Sixteen of the conditions with the CVSD and only six with the LPC-10 are acceptable while six of the CVSD and ten of the LPC-10 are unacceptable, based on the performance criteria.

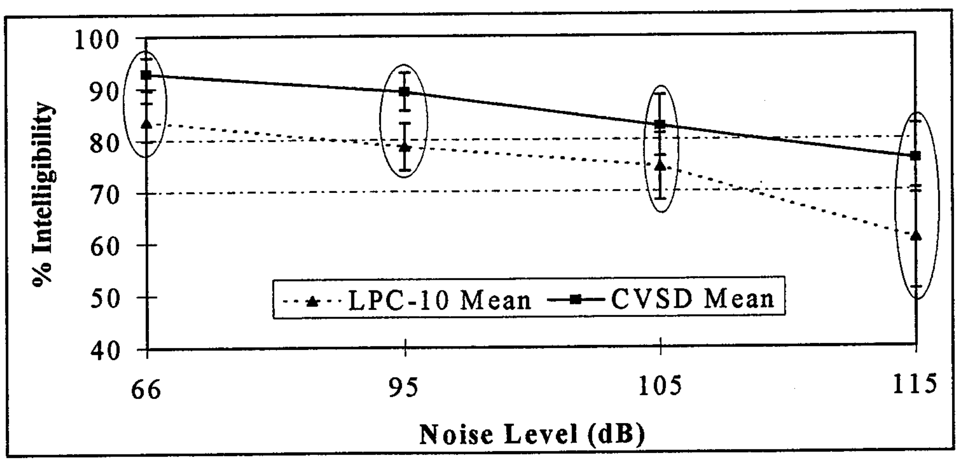



*Figure 31: Phase III - LPC-10 versus CVSD vocoder with the C-130E spectrum, H-157 headset, and female subjects.*

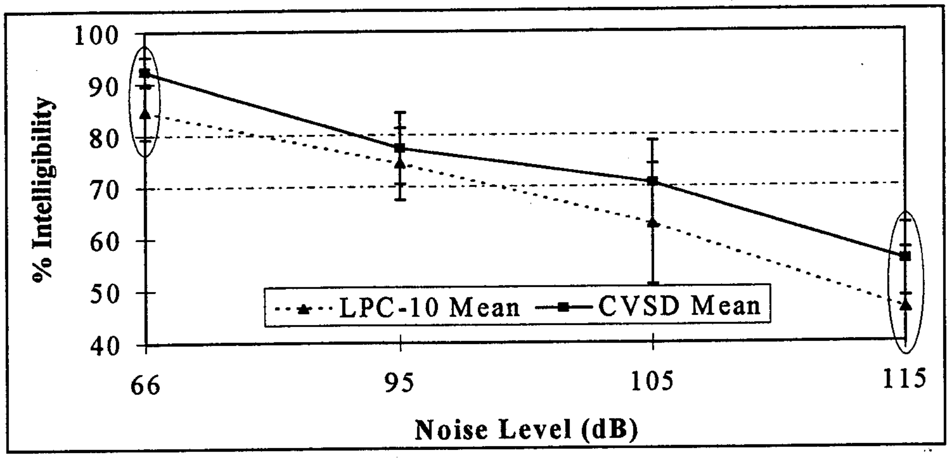



*Figure 32: Phase III - LPC-10 versus CVSD vocoder with the C-141B spectrum, H-157 headset, and female subjects.*

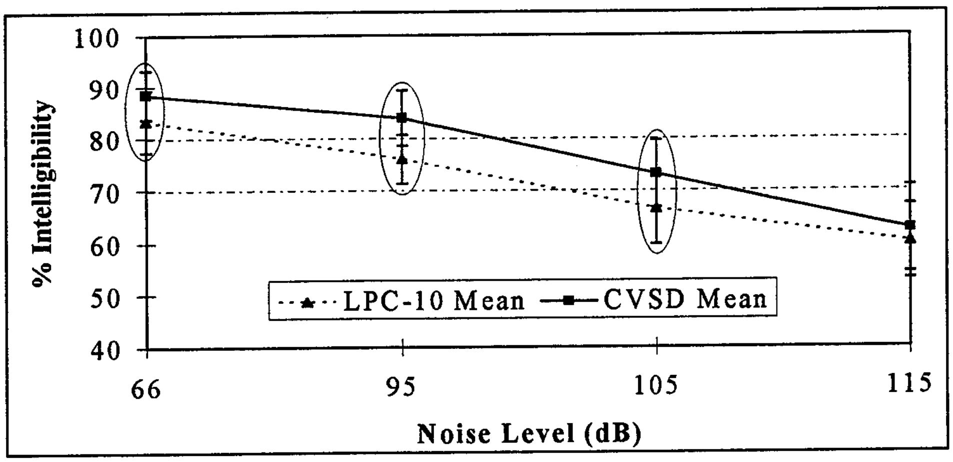



*Figure 33: Phase III - LPC-10 versus CVSD vocoder with the F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and female subjects.*

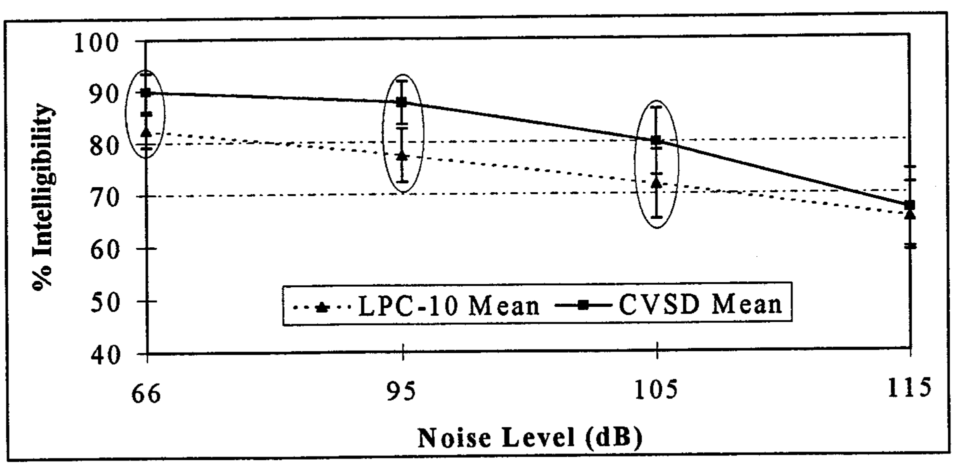



*Figure 34: Phase III - LPC-10 versus CVSD vocoder with the MH-53 helicopter spectrum, SPH-4AF helmet, and female subjects.*

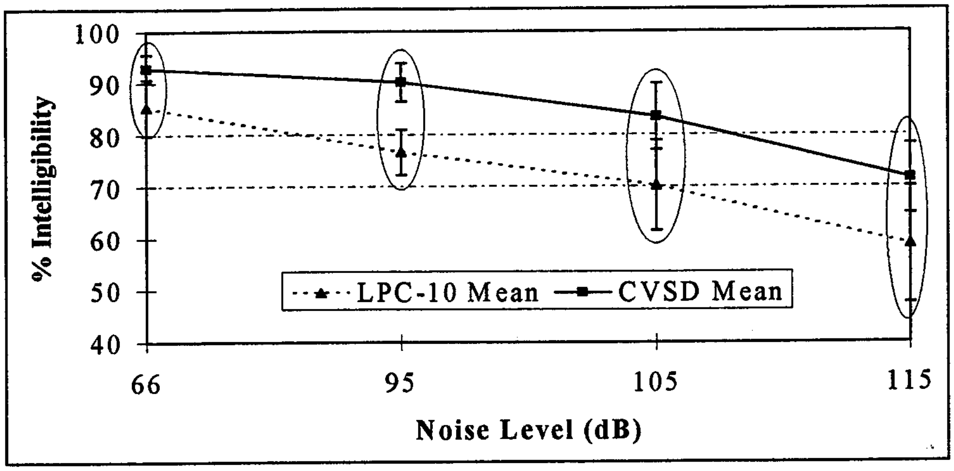



*Figure 35: Phase III - LPC-10 versus CVSD vocoder with the C-130E spectrum, H-157 headset, and male subjects.*

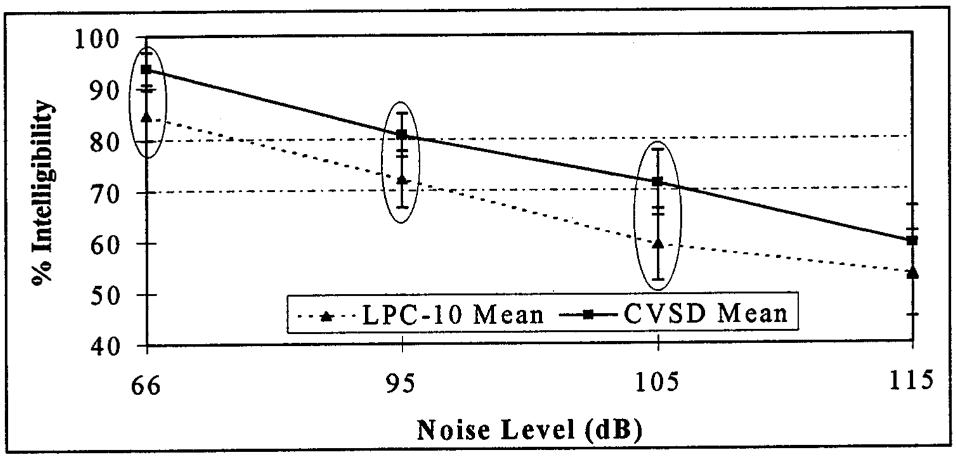



*Figure 36: Phase III - LPC-10 versus CVSD vocoder with the C-141B spectrum, H-157 headset, and male subjects.*

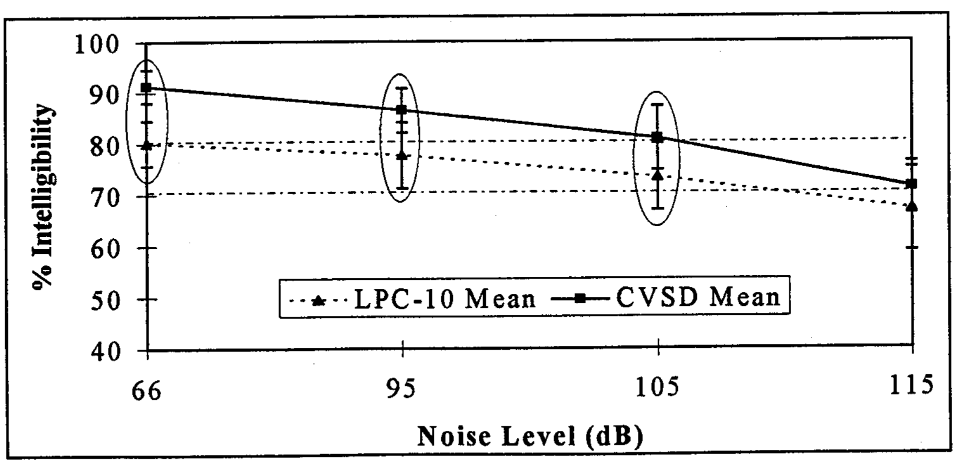



*Figure 37: Phase III - LPC-10 versus CVSD vocoder with the F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and male subjects.*

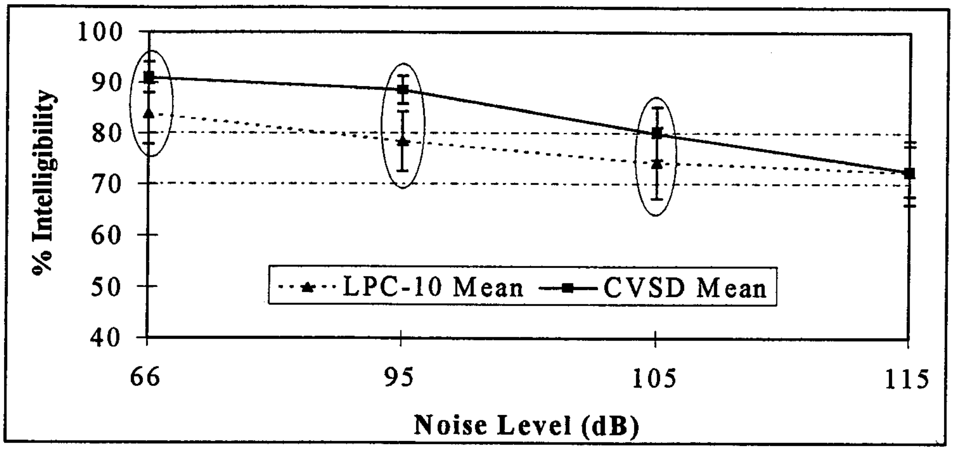



*Figure 38: Phase III - LPC-10 versus CVSD vocoder with the MH-53 helicopter spectrum, SPH-4AF helmet, and male subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 83.7 ± 3.8 | 78.7 ± 4.5 | 74.8 ± 6.5 | 60.7 ± 9.7 |
| CVSD - % avg. intelligibility ± standard deviation | 92.8 ± 3.2 | 89.3 ± 3.7 | 82.7 ± 5.9 | 76.1 ± 6.8 |
| Difference in Means | -9.1 | -10.6 | -7.9 | -15.4 |
| T-score | -5.86 | -5.76 | -2.86 | -4.10 |

*Table 23: Phase III - LPC-10 versus CVSD vocoder with the C-130E spectrum, H-157 headset, and female subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 84.6 ± 5.4 | 74.4 ± 7.0 | 62.8 ± 11.6 | 46.3 ± 11.6 |
| CVSD - % avg. intelligibility ± standard deviation | 92.3 ± 3.2 | 77.4 ± 6.9 | 70.6 ± 8.2 | 55.7 ± 7.1 |
| Difference in Means | -7.7 | -3.0 | -7.9 | -9.3 |
| T-score | -4.02 | -0.95 | -1.75 | -2.17 |

*Table 24: Phase III - LPC-10 versus CVSD vocoder with the C-141B spectrum, H-157 headset, and female subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 83.3 ± 6.0 | 76.0 ± 4.7 | 66.5 ± 6.8 | 60.1 ±7.1 |
| CVSD - % avg. intelligibility ± standard deviation | 88.5 ± 4.7 | 83.9 ± 5.3 | 73.1 ± 6.4 | 62.5 ± 8.2 |
| Difference in Means | -5.2 | -7.9 | -6.6 | -2.4 |
| T-score | -2.13 | -3.53 | -2.22 | -0.68 |

*Table 25: Phase III - LPC-10 versus CVSD vocoder with the F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and female subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 82.3 ± 3.2 | 77.5 ± 5.1 | 71.8 ± 6.6 | 65.4 ± 6.5 |
| CVSD - % avg. intelligibility ± standard deviation | 89.8 ± 3.7 | 87.6 ± 4.1 | 79.9 ± 6.4 | 67.1 ± 7.5 |
| Difference in Means | -7.5 | -10.1 | -8.1 | -1.7 |
| T-score | -4.85 | -4.87 | -2.82 | -0.53 |

*Table 26: Phase III - LPC-10 versus CVSD vocoder with the MH-53 helicopter spectrum, SPH-4AF helmet, and female subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 85.3 ± 5.4 | 76.7 ± 4.4 | 70.2 ± 8.8 | 58.8 ± 11.3 |
| CVSD - % avg. intelligibility ± standard deviation | 92.9 ± 2.7 | 90.1 ± 3.7 | 83.4 ± 6.4 | 71.6 ± 6.7 |
| Difference in Means | -7.6 | -13.4 | -13.2 | -12.8 |
| T-score | -3.99 | -7.40 | -3.85 | -3.09 |

*Table 27: Phase III - LPC-10 versus CVSD vocoder with the C-130E spectrum, H-157 headset, and male subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 84.7 ± 4.8 | 72.3 ± 5.6 | 59.4 ± 7.1 | 53.5 ± 8.4 |
| CVSD - % avg. intelligibility ± standard deviation | 93.8 ± 3.1 | 80.9 ± 4.2 | 71.5 ± 6.4 | 59.6 ± 7.2 |
| Difference in Means | -9.1 | -8.6 | -12.1 | -6.1 |
| T-score | -5.07 | -3.90 | -4.01 | -1.74 |

*Table 28: Phase III - LPC-10 versus CVSD vocoder with the C-141B spectrum, H-157 headset, and male subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 80.1 ± 4.4 | 77.6 ± 6.4 | 73.5 ± 6.5 | 67.1 ± 8.1 |
| CVSD - % avg. intelligibility ± standard deviation | 91.2 ± 3.2 | 86.4 ± 4.3 | 81.0 ± 6.2 | 71.3 ± 5.0 |
| Difference in Means | -11.1 | -8.8 | -7.5 | -4.2 |
| T-score | -6.51 | -3.61 | -2.64 | -1.40 |

*Table 29: Phase III - LPC-10 versus CVSD vocoder with the F-15A spectrum, HGU-55/P helmet with MBU/P oxygen mask, and male subjects.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| LPC-10 - % avg. intelligibility ± standard deviation | 83.9 ± 6.0 | 78.5 ± 5.8 | 74.2 ± 7.0 | 72.4 ± 6.2 |
| CVSD - % avg. intelligibility ± standard deviation | 91.0 ± 3.1 | 88.6 ± 2.7 | 80.0 ± 5.1 | 72.6 ± 5.0 |
| Difference in Means | -7.1 | -10.1 | -5.8 | -0.2 |
| T-score | -3.35 | -4.95 | -2.12 | -0.12 |

*Table 30: Phase III - LPC-10 versus CVSD vocoder with the MH-53 helicopter spectrum, SPH-4AF helmet, and male subjects.*

**Phase IV**

*Automatic Speech Recognition (Voice Control)*

Recognition accuracy of a speech recognition system is generally measured at two levels: at the sentence level and the word level. The percent correct sentences and words are calculated as:

$$\% \text{ Correct} = \frac{\text{Number Correct}}{\text{Total Number}} \times 100\%$$

*ITT VRS-1290 Speaker-Dependent ASR System*

The recognition accuracy of the ITT VRS-1290 is summarized in Figures 39 through 42 and Tables 31 through 34. There are no significant differences between the perception of female and of male speech in any of the sixteen experimental conditions. Sentence recognition accuracy is fairly consistent for all conditions of the C-130E aircraft at the 115 dB level, where it drops slightly. The sentence recognition accuracies are about 10 to 15 percent less than the word recognition accuracies. The ITT word recognition accuracy is relatively resistant to degradation due to increasing levels of the C-130E noise from the ambient to the 115 dB condition. A maximum reduction of 7 percent correct occurred for the word recognition accuracy compared to a 20 percent reduction for the sentence recognition accuracy with a corresponding 50 dB increase in level of the noise.

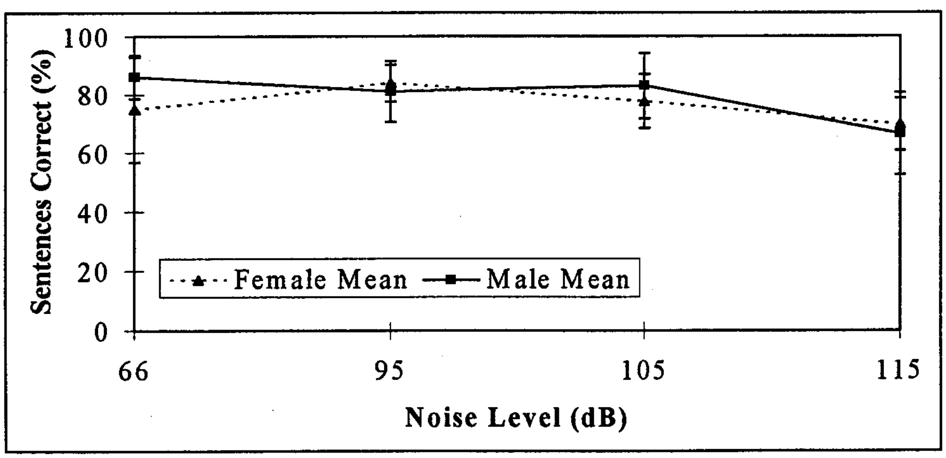



*Figure 39: Phase IV - Male versus female sentence recognition with the ITT ASR system, C-130E spectrum, and the M-87 microphone.*

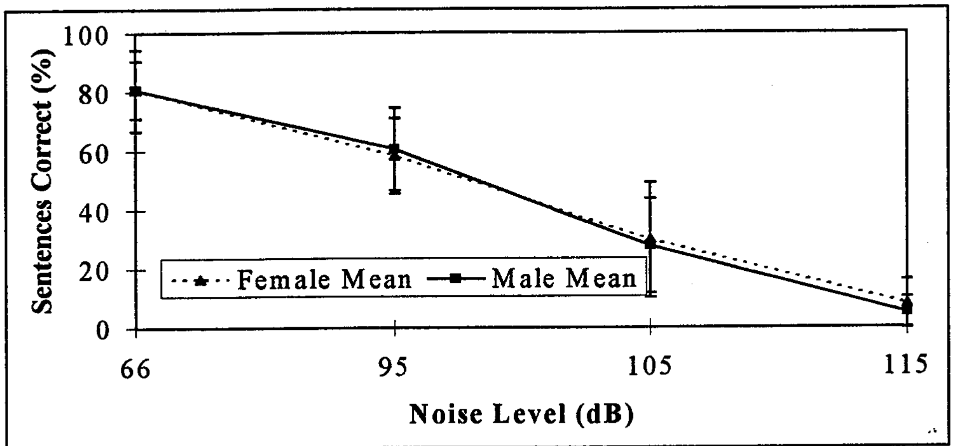



*Figure 40: Phase IV - Male versus female sentence recognition with the ITT ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. sentence recognition ± standard deviation | 74.9 ± 18.0 | 83.9 ± 6.3 | 77.6 ± 9.3 | 69.6 ± 9.0 |
| Male - % avg. sentence recognition ± standard deviation | 86.1 ± 7.4 | 81.1 ± 10.5 | 83.0 ± 11.3 | 66.5 ± 14.0 |
| Difference in Means | -11.2 | 2.8 | -5.4 | 3.1 |
| T-score | -1.81 | 0.71 | -1.17 | 0.58 |

*Table 31: Phase IV - Male versus female sentence recognition with the ITT ASR system, C-130E spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. sentence recognition ± standard deviation | 80.7 ± 13.8 | 58.2 ± 12.7 | 29.7 ± 19.5 | 8.0 ± 8.0 |
| Male - % avg. sentence recognition ± standard deviation | 80.8 ± 9.8 | 60.5 ± 14.0 | 27.6 ± 16.0 | 5.1 ± 5.1 |
| Difference in Means | -0.1 | -2.3 | 2.1 | 2.9 |
| T-score | -0.02 | -0.38 | 0.26 | 0.99 |

*Table 32: Phase IV - Male versus female sentence recognition with the ITT ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

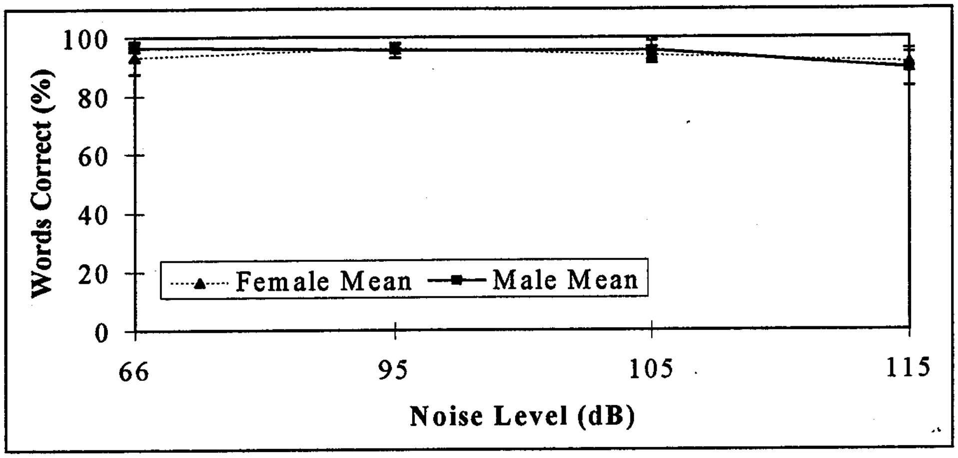



*Figure 41: Phase IV - Male versus female word recognition with the ITT ASR system, C-130E spectrum, and the M-87 microphone.*

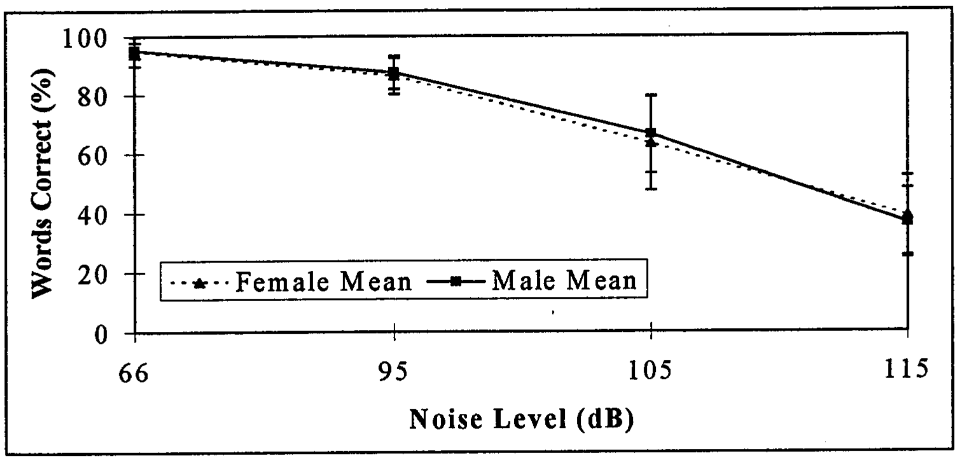



*Figure 42: Phase IV - Male versus female word recognition with the ITT ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. word recognition ± standard deviation | 93.2 ± 5.6 | 96.2 ± 1.7 | 94.0 ± 2.7 | 91.5 ± 3.3 |
| Male - % avg. word recognition ± standard deviation | 96.7 ± 1.6 | 95.5 ± 2.4 | 95.5 ± 3.4 | 89.7 ± 6.4 |
| Difference in Means | -3.5 | 0.7 | -1.5 | 1.8 |
| T-score | -1.93 | 0.73 | -1.13 | 0.80 |

*Table 33: Phase IV - Male versus female word recognition with the ITT ASR system, C-130E spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. word recognition ± standard deviation | 94.6 ± 5.0 | 86.4 ± 6.2 | 63.5 ± 15.9 | 38.9 ± 13.5 |
| Male - % avg. word recognition ± standard deviation | 95.2 ± 2.6 | 87.5 ± 5.7 | 66.5 ± 13.2 | 36.4 ± 11.7 |
| Difference in Means | -0.6 | -1.1 | -3.0 | 2.5 |
| T-score | -0.31 | -0.42 | -0.47 | 0.43 |

*Table 34: Phase IV - Male versus female word recognition with the ITT ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

There are no statistically significant differences between female and male speech in any of the experimental conditions for the ITT ASR system in MH-53 aircraft noise. However, the recognition accuracy of the ITT ASR system is vulnerable to the MH-53 noise spectrum and levels. Recognition accuracies drop sharply as the level of the MH-53 noise increases in increments of 10 dB from 95 to 115 dB. The 29 dB increase in level from ambient (66 dB) to 95 dB is much flatter than it appears on the graph. The 20 dB increase in levels of the noise from 95 to 105 dB is associated with a 50 percent reduction in both words and sentence recognition rate. A 70 percent decrease in word recognition accuracy is observed for the same 20 dB increase in level.

Overall, the ITT ASR system works well in the C-130E noise spectrum and levels; however, it appears to be easily degraded by the MH-53 noise spectrum. This suggests that those working on ASR applications in military aircraft must accomplish more work on ITT ASR performance as a function of spectrum before achieving operational effectiveness. The ITT ASR system is robust in increasing levels of the C-130E noise spectrum, revealing only small reductions in recognition. Projected operational performance is unacceptable in the higher levels of the noise and the recognition accuracy decreases rapidly with the increasing levels of the noise. Both measures of recognition accuracy are adversely affected by the increasing levels of the noise.

*IBM VoiceType ASR System*

Performance data for the IBM VoiceType ASR system are summarized in Figures 43 through 46 and Tables 35 through 38. There are no statistically significant differences between the recognition accuracies of the female and male speech, with the exception of the MH-53 noise spectrum at 115 dB. This exception is attributed, in part, to the improvement in performance of the male speech at 115 dB over that of the 105 dB data. However, female recognition scores were higher than those for male speech in all of the ambient 66 dB noise conditions of both the C-130E and the MH-53 aircraft. The IBM VoiceType is relatively resistant to noise induced reductions in recognition accuracy, displaying somewhat flat performance as a function of the increasing levels of the noise. A maximum reduction of only 14 percent is observed for the word

recognition accuracy and 13 percent for the sentence recognition accuracy over the 50 dB increase in levels of the C-130E noise.

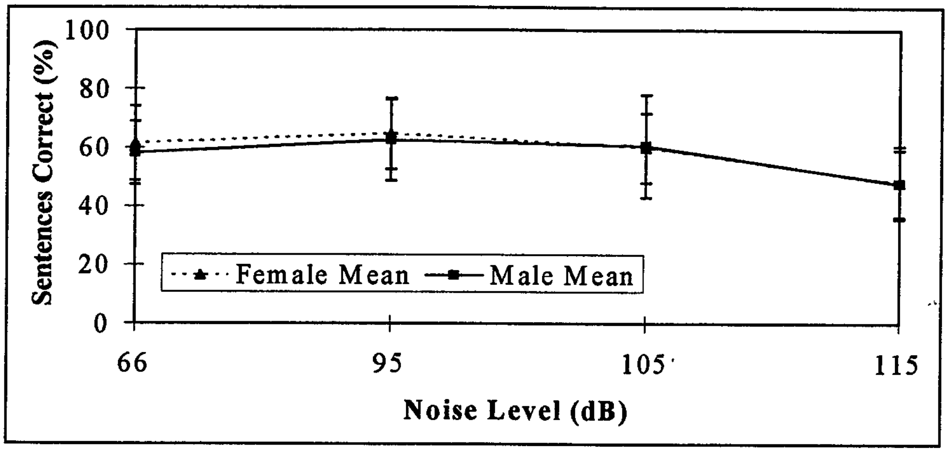



*Figure 43: Phase IV - Male versus female sentence recognition with the IBM ASR system, C-130E spectrum, and the M-87 microphone.*

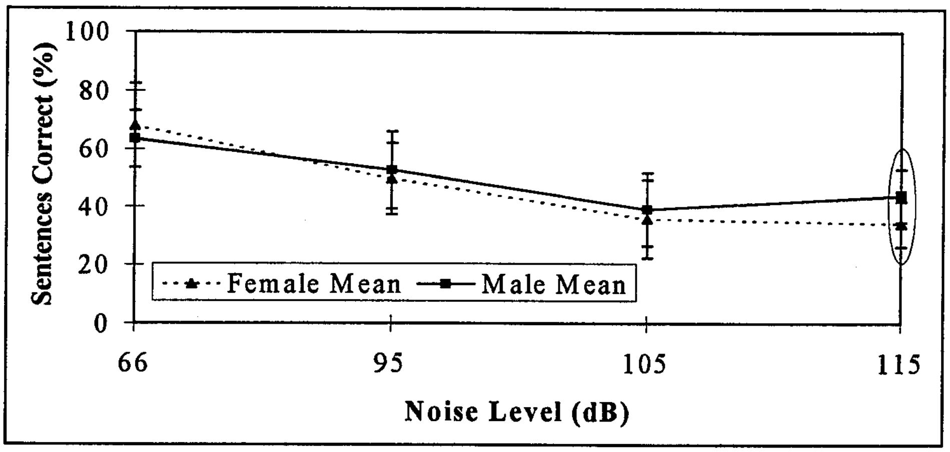



*Figure 44: Phase IV - Male versus female sentence recognition with the IBM ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. sentence recognition ± standard deviation | 61.5 ± 12.7 | 64.8 ± 12.1 | 60.0 ± 11.9 | 48.1 ± 12.5 |
| Male - % avg. sentence recognition ± standard deviation | 58.1 ± 10.8 | 62.5 ± 13.7 | 60.6 ± 17.6 | 47.9 ± 11.5 |
| Difference in Means | 3.4 | 2.3 | -0.6 | 0.2 |
| T-score | 0.64 | 0.39 | -0.09 | 0.04 |

*Table 35: Phase IV - Male versus female sentence recognition with the IBM ASR system, C-130E spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. sentence recognition ± standard deviation | 68.0 ± 14.3 | 49.8 ± 12.4 | 35.8 ± 13.6 | 34.2 ± 7.9 |
| Male - % avg. sentence recognition ± standard deviation | 63.5 ± 9.9 | 52.7 ± 13.3 | 39.1 ± 12.7 | 44.1 ± 9.1 |
| Difference in Means | 4.5 | -2.9 | -3.3 | -9.9 |
| T-score | 0.83 | -0.51 | -0.57 | $\boxed{-2.59}$ |

*Table 36: Phase IV - Male versus female sentence recognition with the IBM ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

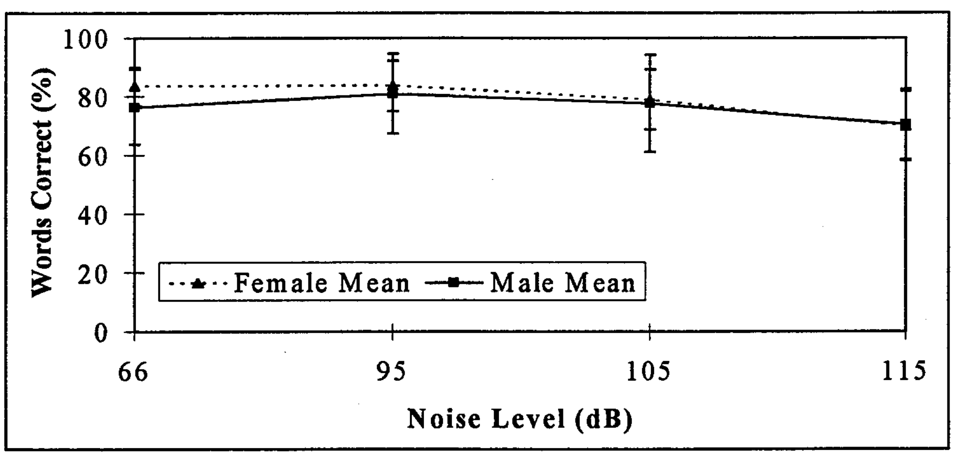



*Figure 45: Phase IV - Male versus female word recognition with the IBM ASR system, C-130E spectrum, and the M-87 microphone.*

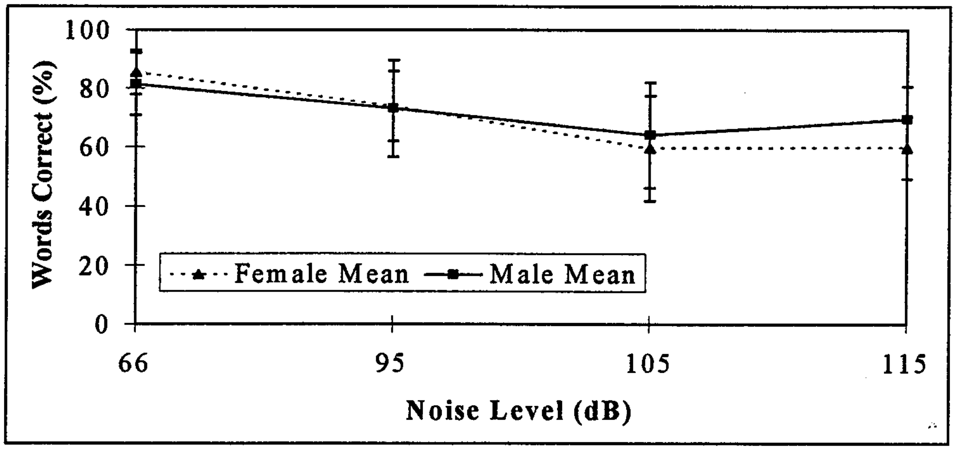



*Figure 46: Phase IV - Male versus female word recognition with the IBM ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. word recognition ± standard deviation | 83.5 ± 6.4 | 83.7 ± 8.7 | 79.0 ± 10.3 | 69.9 ± 11.9 |
| Male - % avg. word recognition ± standard deviation | 76.5 ± 12.8 | 81.1 ± 13.5 | 77.7 ± 16.6 | 70.6 ± 11.9 |
| Difference in Means | 7.0 | 2.6 | 1.3 | -0.7 |
| T-score | 1.56 | 0.52 | 0.22 | -0.12 |

*Table 37: Phase IV - Male versus female word recognition with the IBM ASR system, C-130E spectrum, and the M-87 microphone.*

| | 66 dB | 95 dB | 105 dB | 115 dB |
|---|---|---|---|---|
| Female - % avg. word recognition ± standard deviation | 85.5 ± 7.6 | 73.9 ± 12.0 | 59.7 ± 17.9 | 59.8 ± 10.6 |
| Male - % avg. word recognition ± standard deviation | 81.5 ± 10.7 | 73.1 ± 16.5 | 64.1 ± 17.9 | 69.5 ± 11.2 |
| Difference in Means | 4.0 | 0.8 | -4.4 | -9.7 |
| T-score | 0.97 | 0.12 | -0.55 | -2.00 |

*Table 38: Phase IV - Male versus female word recognition with the IBM ASR system, MH-53 helicopter spectrum, and the M-87 microphone.*

The maximum average score for the IBM system in the MH-53 noises was 85 percent correct in the 66 dB noise. The standard deviations of the male speech were generally larger than those of the female speech when using the IBM ASR system.

Overall, the IBM system exhibits some resistance to degradation by high levels of the two aircraft noises used in this phase of the study, showing lowest scores of around 35 and 40 percent correct. However, it also shows some difficulties by its inability to perform at higher than 85 percent correct in the low level ambient noise condition, which had a relatively flat spectrum and a moderate level of 66 dB. The IBM system operates in both spectra similarly, suggesting its potential for a broader range of applications.

## SUMMARY

Overall results from Phases I, II, and III reveal that the mean percent correct intelligibility of female produced speech was lower than the mean intelligibility of male produced speech by as much as ten percent, and more. This general trend indicated that the amount of the difference between the male and female speech increased as the level of the noise condition increased. The maximum effect usually, but not always, occurred at the condition of highest level of noise. The data also indicated a number of conditions for which the average differences between the female and male speech were statistically significant ($p < 0.05$). These conditions of statistical significance did not follow the trend displayed by the decreasing speech communication effectiveness with increasing level of noise, but were somewhat random in occurrence. However, each one of those conditions verifies female speech is less intelligible than male speech at least 95 percent of the time and that the data demonstrating poorer female speech perception are real. The mean differences for the statistically significant conditions ranged from about three to six percent; however, other conditions with mean differences within this range were not statistically significant. The differences between these averages for both sets of data are relatively small and represent two and three word errors in an MRT list of 50 words. Observation of the percent correct intelligibility data reveals very few situations where a three to six percent difference is meaningful in an operational situation.

Perception of speech in the operational situation was also evaluated using the performance criteria or biocommunications guidelines described earlier. Percent correct intelligibility is compared to benchmark values in the regions below, between, and above the 70 and 80 percent correct intelligibility levels. Laboratory performance exceeding 80 percent correct translates to acceptable operational performance; performance below 70 percent is unacceptable. Performance in the marginal area between the 70 and 80 percent values means operational performance may or may not be acceptable, depending on the specific conditions and requirements. The laboratory values close to 70 and 80 percent (which are not pass-fail values) are in the fringe areas and may require more information than just the intelligibility scores for a confident estimation of the real world performance. The overall speech intelligibility performance for the conditions in Phases I, II, and III are summarized relative to the performance criteria in Table 15. The data are coded such that estimated operational acceptability is equal to + (80 percent and above), marginal acceptability is equal to $\pm$, and operational unacceptability is equal to - (69 percent and below).

| Phase | Aircraft | Vocoder | Gender | Level of Aircraft Noise (dB) 66 | 95 | 105 | 115 |
|---|---|---|---|---|---|---|---|
| I | C-130E | N/A | M | + | + | + | + |
| | | | F | + | + | ± | ± |
| | C-141B | N/A | M | + | + | + | - |
| | | | F | + | + | ± | - |
| | F-15A | N/A | M | + | + | + | ± |
| | | | F | + | + | + | - |
| | MH-53 | N/A | M | + | + | + | ± |
| | | | F | + | + | + | - |
| II | C-130E | N/A | M | + | + | + | + |
| | | | F | + | + | + | + |
| | C-141B | N/A | M | + | + | + | - |
| | | | F | + | + | ± | - |
| | MH-53 | N/A | M | + | + | + | + |
| | | | F | + | + | + | + |
| III | C-130E | LPC-10 | M | + | ± | ± | - |
| | | | F | + | ± | ± | - |
| | | CVSD | M | + | + | + | ± |
| | | | F | + | + | + | ± |
| | C-141B | LPC-10 | M | + | ± | - | - |
| | | | F | + | ± | - | - |
| | | CVSD | M | + | + | ± | - |
| | | | F | + | ± | ± | - |
| | F-15A | LPC-10 | M | + | ± | ± | - |
| | | | F | + | ± | - | - |
| | | CVSD | M | + | + | + | ± |
| | | | F | + | + | ± | - |
| | MH-53 | LPC-10 | M | + | ± | ± | ± |
| | | | F | + | ± | ± | - |
| | | CVSD | M | + | + | + | ± |
| | | | F | + | + | ± | - |

*Table 39: Summary table of average percent correct intelligibility of female and male speech in Phases I and II, evaluated by performance criteria. Acceptable = +, marginal = ±, and unacceptable = -.*

Phase III data indicate female speech is not significantly more vulnerable than male speech to specific vocoders. However, perception of female speech using the DoD standard LPC-10 vocoder is unacceptable in all four aircraft noises at the levels of 105 and 115 dB. Difficulties may be expected in the operational situation for both males and females at these levels.

Based on the results of the speech recognition experiments in Phase IV, there is no significant difference between the speech recognition accuracy of male and female speech in the cockpit noise conditions tested. Neither speech recognition system tested was optimized for performance, in general, and specifically for performance in noise. Therefore conclusions as to the relative performance of these two systems should not be made. Also no conclusions should be made as to the performance in other noise conditions, vocabularies, or in operational conditions. There should be no concern as to the relative performance between male and female speech recognition performance either for speaker-dependent or speaker-independent systems. However, additional work needs to be accomplished to improve overall performance of ASR systems in high noise conditions.

## CONCLUSIONS

1. Mean female speech is less intelligible than mean male speech in all experimental conditions measured in Phases I and II. However, the differences in intelligibility are not always statistically significant, and when statistically significant may not be meaningful in operational situations.

2. These mean differences in intelligibility between male and female speech tend to increase as the levels of the noises increase (from 66 dB to 115 dB).

3. The statistically significant differences between mean intelligibility of male and female speech occurred in a somewhat random fashion. No patterns emerged that were associated with the experimental variables.

4. Examination of the four aircraft cockpit noise spectra at cruise (fixed wing) and hover (rotary wing) indicates that female speech is five to seven percent less intelligible than male speech during cruise. However, both types of speech are acceptable in the C-130E, C-141B, and MH-53. Male speech is marginal and female speech unacceptable in the F-15A noise at the 115 dB level.

5. Female speech is unacceptable in the 115 dB noise of the C-141B, F-15A, and MH-53 and is marginal in the C-130E, 115 dB noise and the F-15A, 105 dB level of noise. Male speech is unacceptable in the 115 dB noise of the C-141B and marginal in the 115 dB noise of the F-15A.

6. Using the M-162 noise-cancelling microphone, both male and female (less than male) speech intelligibility were acceptable in all C-130E and MH-53 noise environments and both were unacceptable in the C-141B spectrum at 115 dB.

7. Speech intelligibility of both female and male speech with the M-162 microphone was as much as 12 percent better than with the M-87 microphone. The greatest improvements occurred in the highest levels of noise.

8. The perception of female speech in noise is not significantly more vulnerable than male speech when processed by the individual LPC-10 and CVSD vocoders. There should be no concern over the relative male-female speech performance for the vocoders studied.

9. The perception of both female and male speech processed by the LPC-10 vocoder is not acceptable for any aircraft noise condition examined. Performance in the ambient condition (66 dB) is in the low range of acceptability.

10. There are no significant differences between the recognition accuracies of male and female speech in the cockpit-noise environments investigated for either ASR system. Although neither ASR system was optimized for performance in noise, there should be no concern for gender differences in recognition accuracy for the systems utilized in the study.

11. Recognition accuracy of the ITT system for both male and female speech is very dependent on noise spectrum. Average word recognition accuracies exceeded 90 percent in all levels of the C-130E noise spectrum. In contrast, word recognition accuracy dropped from 87 percent in the 95 dB level to about 37 percent in the 115 dB level of the helicopter noise.

## RECOMMENDATIONS

Interpretations of the data suggest that the following actions might alleviate the voice communications deficiencies identified in this study. These recommendations can be validated with additional experimentation in the unique voice communications emulation facilities of the Bioacoustics and Biocommunications Branch.

1. Replace the M-87 noise-cancelling microphones with the M-162 noise-cancelling microphones. This would immediately bring the perception of female speech to the current perception level of male speech using the M-87 microphone. The speech intelligibility of the male speech would also experience comparable improvements.

2. Provide headsets and helmets with appropriate active noise reduction (ANR) capability. Because of our extensive experience with ANR technology, we predict that this technology would improve the speech intelligibility in the cockpit environment by reducing noise levels at the eardrum. The Air Force has developed a flight-worthy circumaural headset technology that will undergo Operational Test and Evaluation in the near term.

3. Complete development of a lightweight ANR headset for non-flight-helmet applications such as C-130E and C-141 type aircraft. Because of our experience, we predict that this new technology would also improve communications in these aircraft by reducing noise levels at the eardrum.

4. The LPC-10 vocoder, the DoD standard, is vulnerable to noise at both the talker and listener. Provide a good noise exclusion headset for the listener and an effective microphone noise shield for the talker. Consider speech enhancement systems designed to extract the speech signal from the noise. Additionally, attempt to upgrade the system performance within its current

constraints (i.e. bandwidth) to obtain a more acceptable level of performance in noise for both male and female speech.

5. Recognition accuracy varies widely among ASR systems. Systems should be evaluated in the environment (i.e., noise, vibration, heat) in which they will be used prior to their acquisition and installation. Additional work is required to improve overall performance of ASR systems in high level noise conditions.

## POST LOG

Overall, the perception of both female and male speech is degraded by noise. The amount of the degradation increases as the relative levels of the noise increase. Degradation of this speech is also dependent of the spectrum of the noise and is most affected by the quantity of the noise that is present in the speech frequency regions. With analog systems, female speech is usually more vulnerable in noise than male speech. Many of these differences are statistically significant and while some are important, others are sufficiently small as to be immaterial in practical operational situations. Differences between female and male speech are insignificant when processed by many digital systems such as automatic speech recognition and speech coding-decoding systems. Both analog and digital systems are components of current AF and DoD audio communications. Some of the communications improvements needed with the systems identified in this study are within the state-of-the-art, whereas others will require new knowledge and advanced technology.

## REFERENCES


1. ANSI S12.6-1984 (R 1990), American National Standard Method for the Measurement of Real-Ear Attenuation of Hearing Protectors.

2. ANSI S3.2-1989 (ASA 85), American National Standard Method for Measuring the Intelligibility of Speech over Communications Systems.

3. Backs, R. W., and Walrath, L. C., "(A) Heart Rate and Auditory Workload During Noise Stress," Proceedings of 6th International Symposium on Aviation Psychology, Vol 2 (A92-44901-19-53), Ohio State University, Columbus OH, 1991.

4. Box, George E. P., Hunter, William G., and Hunter, J. Stuart, Statistics for Experimenters, John Wiley & Sons, Inc., 1978.

5. C-130E In-Flight Crew Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 38, Wright-Patterson AFB OH, September 1975.

6. C-141B In-Flight Crew/Passenger Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 126, Wright-Patterson AFB OH, June 1980.

7. F-15A In-Flight Crew Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 127, Wright-Patterson AFB OH, August 1979.

8. Fletcher, Harvey, Speech and Hearing in Communication, D. Van Nostrand Company, Inc., 1953.

9. Freedman, Jay and Rumbaugh, William A. "Accuracy and Speech of Response to Different Voice Types in A Cockpit Voice Warning System," Air Force Institute of Technology, LSSR 89-83, Wright-Patterson AFB OH, 1983.

10. HH-53C In-Flight Crew Noise, USAF Bioenvironmental Noise Data Handbook, AMRL-TR-75-50, Vol 51, Wright-Patterson AFB OH, October 1975.

11. Holden, James M. and Vensko, George, Voice R&D Progress in the Helicopter Environment, Proceedings of Speech Tech 89, Media Dimensions, New York NY, 1989.

12. House, A. S., Williams, Carl E., Hecker, Michael H. L., and Kryter, Karl D., "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set," Journal of the Acoustical Society of America, 37 (1965), 158-166.

13. McKinley, Richard L., "Voice Communication Research and Evaluation System," Aerospace Medical Research Laboratory, AMRL-TR-80-25, Wright-Patterson AFB OH, 1980.

14. McKinley, Richard L., Nixon, Charles W., and Moore, Thomas J., "Voice Communication Capability of Selected In-Flight Headgear Devices," AGARD Proceedings, Soesterberg, Netherlands, March-April 1981.

15. Moore, Thomas J., Nixon, Charles W., and McKinley, Richard L., "Comparative Intelligibility of Speech Materials Processed by Standard Air Force Communications Systems in the Presence of Simulated Cockpit Noise," AGARD Conference Proceedings No. 311, Aural Communication in Aviation, June 1981.

16. Nixon, Charles W., Moore, Thomas, J., and McKinley, Richard L., "Increase in Jammed Word Intelligibility Due to Training of Listeners," Journal of Aviation and Space Medicine, AAFMRL-81-64, 1982.

17. Nixon, Charles W. and McKinley, Richard L., "Intelligibility in Noise of Three LPC Voice Channels with ANR Headsets," AAMRL-TR-88-063, Wright-Patterson AFB OH 45433, November 1988.

18. Nixon, C. W. and McKinley, R. L., "LPC-10 Intelligibility of Oxygen Masks and Microphones in Noise," Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB OH 45433, AAMRL-TR-88-048, November 1988.

19. Nixon, C. W., Stephenson, M., McKinley, R. L., Fisher, V., and Jacobs, M., "Intelligibility of Digital Speech Masked by Noise: Normal Hearing and Hearing Impaired Listeners," Aerospace Medical Research Laboratory, AAMRL-TR-90-1990, Wright-Patterson AFB OH 45433, June 1990.

20. Nixon, Charles W., "Voice Communications and Positive Pressure Breathing," AFAMRL-TR-84-009, Wright-Patterson AFB OH, 45433, January 1984.

21. Prinzo, O. Veronika and Britton, Thomas W., "ATC/Pilot Voice Communications - A Survey of the Literature," Civil Aeromedical Institute, FAA, Oklahoma City OK 73125, (p. 6), November 1993.

22. Simpson, Carol, "Evaluation of Speech Recognizers for use in Advanced Combat Helicopter Crew Station Research and Development," Ames Research Center, NASA Contractor Report 177547, US Army Aviation Systems Command, TM-90-A-001, Moffet Field CA 94035, March 1990.

23. Simpson, Carol A., Marchionda-Frost, Kristine, and Navarro, Teresa, "Comparison of Voice Types for Helicopter Voice Warning Systems, NASA Contract NAS2-11341, Report 841611, Ames Research Center, Moffett Field CA.

24. Williamson, David T., and Curry, David G., Speech Recognition Performance Evaluation in Simulated Cockpit Noise," Proceedings of Speech Tech 84, Media Dimensions, New York NY, 1984.

25. Werkowitz, Eric, "Speech Recognition in the Tactical Environment: The AFTI-F-16 Voice Command," Proceedings of Speech Tech 84, Media Dimensions, New York NY, 1984.

# APPENDIX A

| Microphone Type | Mic Number | Frequency (Hz) 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
|---|---|---|---|---|---|---|---|---|
| Brüel & Kjær | 4134 | 1.60 | 0.40 | 0.67 | 0.45 | 1.30 | 0.81 | 1.06 |
| M-87 | 1 | 2.12 | 4.00 | 7.80 | 8.97 | 9.63 | 5.23 | 1.23 |
| M-87 | 2 | 1.43 | 2.71 | 6.36 | 9.31 | 7.29 | 4.08 | 1.22 |
| M-87 | 3 | 1.63 | 3.60 | 8.46 | 9.96 | 7.60 | 4.78 | 1.09 |
| M-87 | 4 | 1.46 | 2.99 | 9.00 | 6.76 | 4.93 | 5.30 | 0.78 |
| M-87 | 5 | 2.14 | 4.32 | 9.09 | 8.68 | 9.72 | 3.20 | 1.58 |
| M-87 | 6 | 1.63 | 3.11 | 6.69 | 11.52 | 13.70 | 4.10 | 1.74 |
| M-87 | 7 | 1.20 | 2.14 | 4.63 | 9.68 | 5.89 | 5.38 | 2.00 |
| M-87 | 8 | 2.47 | 4.96 | 9.46 | 10.75 | 11.93 | 5.11 | 1.29 |
| M-87 | 9 | 2.47 | 4.83 | 9.55 | 8.97 | 9.44 | 5.14 | 1.64 |
| M-87 | 10 | 1.52 | 1.22 | 6.33 | 9.98 | 9.97 | 4.74 | 2.53 |

*Table 40: M-87 microphone calibration data in volts rms.*

| Microphone Type | Mic Number | Frequency (Hz) 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
|---|---|---|---|---|---|---|---|---|
| Brüel & Kjær | 4134 | 1.60 | 0.39 | 0.66 | 0.44 | 1.30 | 0.80 | 1.05 |
| M-162 | 11 | 0.83 | 0.86 | 0.85 | 0.81 | 0.87 | 0.56 | 0.31 |
| M-162 | 12 | 0.69 | 0.71 | 0.71 | 0.65 | 0.72 | 0.49 | 0.44 |
| M-162 | 13 | 0.96 | 0.87 | 0.90 | 0.75 | 0.90 | 0.56 | 0.31 |
| M-162 | 14 | 0.78 | 0.69 | 0.71 | 0.61 | 0.70 | 0.40 | 0.37 |
| M-162 | 15 | 0.68 | 0.73 | 0.77 | 0.68 | 0.85 | 0.58 | 0.39 |
| M-162 | 16 | 0.70 | 0.65 | 0.70 | 0.63 | 0.77 | 0.80 | 0.27 |
| M-162 | 17 | 0.69 | 0.62 | 0.66 | 0.57 | 0.69 | 0.41 | 0.28 |
| M-162 | 18 | 0.63 | 0.55 | 0.89 | 0.49 | 0.58 | 0.34 | 0.40 |
| M-162 | 19 | 0.96 | 0.89 | 0.94 | 0.81 | 1.00 | 0.62 | 0.30 |
| M-162 | 20 | 0.86 | 0.81 | 0.88 | 0.78 | 0.98 | 0.64 | 0.29 |

*Table 41: M-162 microphone calibration data in volts rms.*

| Microphone Type | Mic Number | Frequency (Hz) 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
|---|---|---|---|---|---|---|---|---|
| Brüel & Kjær | 4134 | 1.60 | 0.39 | 0.66 | 0.44 | 1.30 | 0.80 | 1.05 |
| M-169 | 21 | 2.28 | 4.26 | 8.07 | 11.67 | 12.11 | 10.01 | 3.87 |
| M-169 | 22 | 2.15 | 3.99 | 7.25 | 10.75 | 11.73 | 10.41 | 5.76 |
| M-169 | 23 | 2.35 | 4.35 | 8.00 | 11.76 | 12.00 | 10.28 | 5.56 |
| M-169 | 24 | 1.96 | 3.61 | 6.85 | 10.79 | 11.86 | 10.23 | 6.26 |
| M-169 | 25 | 1.97 | 3.60 | 6.50 | 9.40 | 10.47 | 9.76 | 5.73 |
| M-169 | 26 | 1.88 | 3.37 | 6.25 | 9.32 | 10.32 | 10.40 | 5.88 |
| M-169 | 27 | 2.12 | 3.94 | 7.46 | 11.34 | 11.91 | 9.87 | 5.27 |
| M-169 | 28 | 2.09 | 3.91 | 7.24 | 9.79 | 10.50 | 8.71 | 6.10 |
| M-169 | 29 | 2.10 | 3.66 | 6.03 | 8.13 | 9.54 | 9.45 | 4.48 |
| M-169 | 30 | 2.03 | 3.70 | 6.58 | 9.93 | 11.42 | 10.63 | 6.66 |

*Table 42: M-169 microphone calibration data in volts rms.*

# APPENDIX B

| | | Frequency (Hz) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Headset/ Helmet Type | Average (m) or one standard deviation (1σ) | 125 | 250 | 500 | 1000 | 2000 | 3150 | 4000 | 6300 | 8000 |
| HGU-55/P | m | 8 | 2 | 10 | 23 | 37 | 42 | 45 | 47 | 47 |
| | 1σ | 4.3 | 4.0 | 4.6 | 5.3 | 5.8 | 4.8 | 5.2 | 6.9 | 7.1 |
| H-157A | m | 10 | 12 | 18 | 32 | 38 | 39 | 37 | 37 | 35 |
| | 1σ | 2.6 | 2.9 | 3.6 | 6.2 | 4.3 | 4.9 | 6.1 | 7.3 | 6.0 |
| SPH-4AF | m | 14 | 13 | 24 | 37 | 38 | 40 | 40 | 45 | 43 |
| | 1σ | 2.7 | 2.2 | 2.2 | 5.3 | 2.6 | 4.1 | 4.3 | 5.0 | 4.8 |

*Table 43: Average and standard deviation of headset/helmet attenuation (sound pressure level, dB).*

# APPENDIX C

## Modified Rhyme Test (MRT) Sample Score Sheet

| | | | |
|---|---|---|---|
| 1 | hill | kill | bill |
| | will | till | fill |
| 2 | dim | dill | din |
| | did | dip | dig |
| 3 | fizz | fib | fill |
| | fig | fit | fin |
| 4 | heal | heat | heave |
| | heap | heath | hear |
| 5 | rust | dust | gust |
| | must | just | bust |
| 6 | run | bun | fun |
| | nun | gun | sun |
| 7 | look | shook | book |
| | cook | took | hook |
| 8 | rang | sang | hang |
| | fang | gang | bang |
| 9 | dub | dun | dung |
| | dud | dug | duck |
| 10 | tin | pin | fin |
| | win | sin | din |
| 11 | lake | late | lane |
| | lace | lame | lay |
| 12 | heel | keel | peel |
| | reel | eel | feel |
| 13 | peace | peak | peat |
| | peach | peas | peal |
| 14 | seethe | seek | seem |
| | seed | seep | seen |
| 15 | hold | fold | sold |
| | cold | told | gold |
| 16 | beat | meat | neat |
| | heat | seat | feat |
| 17 | sub | sud | sup |
| | sung | sum | sun |
| 18 | say | day | fay |
| | may | gay | way |
| 19 | race | raze | rave |
| | ray | rake | rate |
| 20 | cane | case | cave |
| | cape | came | cake |
| 21 | kick | wick | tick |
| | pick | lick | sick |
| 22 | back | bass | bath |
| | ban | bat | bad |
| 23 | dip | tip | hip |
| | lip | sip | rip |
| 24 | pale | gale | sale |
| | bale | male | tale |
| 25 | bit | wit | hit |
| | kit | sit | fit |
| 26 | pit | pill | pig |
| | pick | pip | pin |

| | | | |
|---|---|---|---|
| 27 | red | shed | bed |
| | fed | led | wed |
| 28 | cuss | cup | cuff |
| | cut | cud | cub |
| 29 | tease | teach | teal |
| | team | tear | teak |
| 30 | dent | went | sent |
| | rent | tent | bent |
| 31 | pass | pat | pan |
| | pad | path | pack |
| 32 | kid | kit | kin |
| | kill | king | kick |
| 33 | den | ten | hen |
| | then | pen | men |
| 34 | bun | bug | buff |
| | buck | but | bus |
| 35 | pay | pave | pale |
| | pane | pace | page |
| 36 | boil | soil | toil |
| | oil | foil | coil |
| 37 | sap | sad | sag |
| | sass | sat | sack |
| 38 | tame | came | game |
| | fame | name | same |
| 39 | not | hot | tot |
| | pot | lot | got |
| 40 | rest | test | west |
| | best | vest | nest |
| 41 | safe | same | save |
| | sake | sane | sale |
| 42 | sit | sing | sill |
| | sin | sick | sip |
| 43 | big | rig | wig |
| | dig | fig | pig |
| 44 | beam | bean | beat |
| | beak | beach | bead |
| 45 | puff | pub | pup |
| | pun | pus | puck |
| 46 | dark | bark | mark |
| | park | lark | hark |
| 47 | tab | tan | tang |
| | tack | tap | tam |
| 48 | paw | saw | thaw |
| | law | jaw | raw |
| 49 | math | man | map |
| | mass | mat | mad |
| 50 | cop | mop | hop |
| | top | shop | pop |

# APPENDIX D

## Phase IV Automatic Speech Recognition (ASR) Vocabulary

| | | |
|---|---|---|
| Add-new | Go-to | Range |
| after | ground-map | sector-up |
| beacon | heading-up | seven |
| before | I-D-S | Show |
| Change | layer | six |
| comm | minutes | South |
| degrees | Modify | T-A |
| Delete | N-R-P | T-F |
| Display | nine | ten |
| East | North | three |
| eight | north-up | to |
| eighty | one | track-up |
| five | one-sixty | twenty |
| flight-director | page | two |
| flight-plan | pencil-beam | two-forty |
| forty | point | weather |
| four | radar | West |
| Give-me | radar-mode | zero |

# TABLE OF FIGURES

# TABLE OF TABLES